DOCUMENT RESUME

ED 215 905 SE 037 367

AUTHOR	Blosser, Patricia E.; Mayer, Victor J.
TITLE	Investigations in Science Education, Volume 8, Number 1, 1982.
INSTITUTION	ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, Ohio.
PUB DATE	82
NOTE	82p.
AVAILABLE FROM	Information Reference Center (ERIC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 ($1.75 single copy, $6.00 per year subscription price).
JOURNAL CIT	Investigations in Science Education; v8 n1 1982
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Attitudes; *Cognitive Development; Educational Research; Elementary Secondary Education; Higher Education; Methods Courses; *Preservice Teacher Education; Science Curriculum; *Science Education; Science Instruction; Teacher Behavior
IDENTIFIERS	*Science Education Research

ABSTRACT

This issue of Investigations in Science Education contains articles about attitude research in science education. It contains the critiques of nine articles about attitude research as well as three responses to critiques. One response relates to a critique published in an earlier issue while the other two responses are paired with the critiques that provoked the response. One critique-response pair relates to attitude research; the other pair, to research related to cognitive development. Topics related to attitude research vary and include attitude assessment as well as studies of the effects of attitudes of students on instruction in science. (PEB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED215905

INVESTIGATIONS IN SCIENCE EDUCATION

Editor

Patricia E. Blosser
The Ohio State University

Associate Editor

Victor J. Mayer
The Ohio State University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Advisory Board

Willard J. Jacobson (1982)
Teachers College

Gerald Neufeld (1984)
Brandon University

Anton E. Lawson (1984)
Arizona State University

Lowell J. Bethel
University of Texas (1984)

Robert L. Steiner (1983)
University of Puget Sound

National Association for Research in Science Teaching

ERIC® Clearinghouse for Science, Mathematics
and Environmental Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert W. Howe

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 8, Number 1, 1982

Published Quarterly by

The Center for Science and Mathematics Education
College of Education
The Ohio State University
1945 North High Street
Columbus, OH 43210

Subscription Price: $6.00 per year. Single Copy Price: $1.75
Add 50¢ for Canadian mailings and $1.00 for foreign mailings.

SE 037 367

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 8, Number 1, 1982

NOTES FROM THE EDITOR iii

ATTITUDES 1

Moyer, Richard H. "Environmental Attitude Assessment: Another Approach." Science Education, 61(3): 347-356, 1977.
Abstracted by RONALD D. SIMPSON 3

Crawley, Frank E. and John W. Shrum. "Effects of Learning Structure Condition on Change in Preference for Science Courses." Journal of Research in Science Teaching, 14(3): 257-262, May 1977.
Abstracted by RICHARD M. SCHLENKER. 13

Fraser, Barry J. "Selection and Validation of Attitude Scales for Curriculum Evaluation." Science Education, 61(3): 317-329, 1977.
Abstracted by JOHN E. PENICK. 20

Novick, S. Shimshon and D. Duvdvani. "The Relationship Between School and Student Variables and the Attitudes Toward Science of Tenth-Grade Students in Israel." Journal of Research in Science Teaching, 13(3): 259-265, 1976.
Abstracted by MICHAEL SZABO 27

Kauchak, D. P. "The Effects of Essay Writing on the Attitudes of Undergraduate Science Methods Students." Journal of Research in Science Teaching, 14(2): 139-143, 1977.
Abstracted by MICHAEL SZABO 33

Moore, Kenneth D. and Shirley Robards. "Field Centered Preservice Elementary Science and Content Reading Methods." Piper, M. and K. D. Moore, (Eds.) Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, 1977.
Abstracted by MICHAEL PADILLA 40

DeBruin, Jerome. "The Effect of a Field-Based Elementary Science Teacher Education Program on Undergraduates' Attitudes Toward Science and Science Teaching," in Piper, M. and K. Moore, (Eds.). Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, Ohio State University, 1977.
Abstracted by DAVID P. BUTTS 44

Jaus, Harold. "An Analysis of the Relationship of Preservice Elementary Teachers' Attitudes Toward Teaching Science and Their Science Teaching Planning Practices," in Piper, M. and K. Moore (Eds.), Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, Ohio State University, 1977.
Abstracted by DAVID P. BUTTS. 47

CRITIQUES AND RESPONSES 49

Berger, C. F. "Investigation of Teacher Behavior: Interaction with New Curriculum Materials." Attitudes Toward Science: Investigations. Piper, M. and K. Moore, (Eds.) Columbus, OH: SMEAC Information Reference Center, 1977.
Abstracted by F. GERALD DILLASHAW and JAMES R. OKEY 51

Berger, C. F. "Investigation of Teacher Behavior: Interaction with New Curriculum Materials," by F. Gerald Dillashaw and James R. Okey. Investigations in Science Education, 8(1): 49-54, 1982. 57

Renner, J. W. "The Relationships Between Intellectual Development and Written Responses to Science Questions." Journal of Research in Science Teaching, 16(4): 279-299, 1979.
Abstracted by M. E. MILLER and M. C. LINN. 60

Renner, John W. "The Relationship Between Intellectual Development and Written Responses to Science Questions" by M. E. Miller and M. C. Linn. Investigations in Science Education, 8(1): 60-68, 1982. 69

Sunal, Dennis. "Analysis of Research on the Educational Uses of A Planetarium," by G. H. Krockover. Investigations in Science Education, 7(1): 57-59, 1981. 78

NOTES FROM THE EDITOR

Regular readers of ISE will notice a change in the format. In these days of increasing costs and stable or decreasing budgets, we have switched from the smaller, perfect-bound format to a less expensive one. There is another, perhaps less-noticeable change--an increased number of pages. When we completed assembling copy for Volume 7, we still had 328 pages of typed copy for use in producing Volume 8. To accommodate abstractors who were eager to see their work in print, we have increased the page count from 66 to 80.

Volume 8, Number 1 of Investigations in Science Education contains analyses of articles focused primarily on attitudes. Attitude research continues to interest many science educators even if their work does not result in findings at a level or statistical significance. Articles in this issue contain descriptions on attitude assessment (Moyer, Fraser) as well as studies of the effects of attitudes of students on instruction (Crawley and Shrum, Novick and Duvdvani, Kauchak, Moore and Robards, DeBruin, Jaus, and Berger).

In the "Critiques and Responses" section of this issue the reader will find two paired critiques and responses. One of these pairings relates to an attitude article; the other, to an article on the assessment of intellectual development. Also included in this section is the response by Sunal to an article critiqued in Volume 7, Number 1. We hope this immediate pairing of analysis/critique and response will benefit those science educators using ISE in college research classes.

Patricia E. Blosser
Editor

Victor J. Mayer
Associate Editor

ATTITUDES

Moyer, Richard H. "Environmental Attitude Assessment: Another Approach."
Science Education, 61(3): 347-356, 1977.
Descriptors--Affective Behavior; *Attitudes; Educational Research; Elementary School Science; Elementary Secondary Education; *Environmental Education; Science Education; *Secondary School Science; *Tests; *Test Validity

Expanded abstract and analysis prepared especially for I.S.E. by Ronald D. Simpson, North Carolina State University.

Purpose

This study was initiated to develop and standardize the Mayor Unobstrusive Survey of Environmental Attitudes (MUSEA). The purpose of MUSEA is to assess feelings of respondents that they might not otherwise divulge. The instrument utilizes projective techniques designed to lead respondents into reacting toward three environmental themes: pollution, population, and ecological relationships.

Rationale

The investigator writes that dozens of attitude instruments have been published during the last decade. Many of these instruments have purported to measure attitudes of students toward various environmental issues. He states further, however, that all of the instruments he has evaluated have been straightforward questionnaires, usually utilizing Likert-type formats. The problem with questionnaires like these, he asserts, is that they are an index of what respondents are willing to say concerning their attitudes; and he cites work by Corey (1937) and Oppenheim (1966) that he says confirms that what people say their attitudes are and what their attitudes really are may not be the same.

It was the investigator's attempt to develop and standardize projective methods of attitude assessment that served as a basic rationale for this study. Indeed, few studies of this kind have been conducted recently that attempt to compare standard methods of attitude assessment with those

The MUSEA is comprised of three themes: Pollution, Population and Ecological Relationships. Subsequently, there are three subscales: Word Association Scale (WAS), Free Choice Scale (FCS), and the Sentence Completion Scale (SCS). The WAS and SCS subscales contain three items relating to each of the three themes. The FCS includes nine items relating to the environment in general. The following information from Table II clarifies the distribution of the 27 items in MUSEA:

	Pollution	Population	Ecological Relationships	Total
Word Association Scale	3	3	3	9
Sentence Completion Scale	3	3	3	9
Free Choice Scale	-	-	-	9
				27

Word association is based on the assumption that responding rapidly to stimuli will lead to individuals revealing information about their feelings that they otherwise may be unwilling to divulge. Also, it is assumed that when subjects have time to ponder their responses that they may rationalize what a "good" or "acceptable" answer should be. The Word Association Scale used in this study is comprised of nine key words that are read to students. Each student is asked to respond as promptly as possible with the first three words that come to mind. In addition to the nine key words, neutral words are randomly distributed to avoid potentially developing mental sets.

The Sentence Completion Scale (SCS) is composed of nine sentence fragments, three for each of the three themes of MUSEA (pollution, population, and ecological relationships). The investigator states that the fragments are worded in the third person so that respondents will not feel as though they are being directly questioned. Oppenheim (1966) is cited as a reference which suggests that more insightful responses are elicited from subjects when this method is used. The assumptions of the SCS are similar to those of the WAS; however, the structure of sentence completion, according to the investigator, may yield more of a more projective format. For this reason, the investigation being analyzed here represents a new direction in attitude research--a direction that should prove to be informative to many science and environmental educators.

The investigator posed five additional questions as he outlined the objectives of this study. The following questions were studied using correlation and multiple regression ANOVA techniques (Kerlinger and Elazar, 1954):

(1) In order to determine if the scales of the MUSEA measure the same attitudes, the following question was formulated: Is there a difference in student scores between scales of the MUSEA?

(2) To assess any possible relationships between the three scales of the MUSEA, pollution, population, and ecological relationships, the following question was asked: Is there a difference in student scores between themes of the MUSEA?

(3) To determine if the MUSEA can be used effectively with urban and rural subjects, the following question was formulated: Is there a significant difference in scores on the MUSEA for urban or rural students?

(4) To study the effectiveness of the MUSEA with subjects living in a small, medium, or large size communities, the following question was investigated: Is there a significant difference in scores on the MUSEA for students from small, medium, or large communities?

(5) To assess whether the MUSEA can be used with males and with females: Is there a significant difference in scores on the MUSEA between male and female subjects?

Research Design and Procedure

The sample for this study included 379 seventh grade students in Colorado. The sample was stratified with respect to community size and setting (rural or urban) and utilized 14 intact classes.

information and be interpreted more easily since it "...cuts down the multiplicity of associations evoked by a single word..." (Sacks and Levy, 1950).

The Free Choice Scale (FCS) consists of nine topics related to the three MUSEA themes. These nine topics are randomly mixed with nine other topics that do not relate to the environment. The 18 topics represent simulated news stories and respondents are asked to select 9 of the 18 for showing on a fictitious news program (students are asked to play the role of a weekly television news program editor). The frequency with which the subject chooses the environmentally-related topics becomes an index of attitude toward environment.

Each item of the WAS and SCS is judged as positive (+), neutral (0), or negative (-). In the FCS, each environmental issue chosen is scored as positive (+). One point is assigned a positive response and one point subtracted for a negative response. Therefore, the total possible score for each subscale is 9 and for the MUSEA the maximum score possible is 27.

The investigator reports techniques used to estimate validity and reliability. Also, portions of the MUSEA were checked for readability using the Fry Readability Formula.

Descriptive data were compiled for each stratum in the sample, for each theme and subscale, and for total MUSEA scores. Multiple regression analyses were used in an attempt to answer the research questions posed and to standardize the MUSEA.

Findings

The mean score for the total sample on the MUSEA was found to be 10.21 (range possible was -18 to + 27) with a standard deviation of 4.66. The following table was used by the investigator to present normative data.

Norming Data for MUSEA

Scale	Mean	Standard Deviation
WAS - pollution	0.01	1.07
WAS - population	0.02	1.48
WAS - ecological relationships	0.70	0.77
WAS - total	2.53	1.86
FCS - total	4.84	1.82
SCS - pollution	0.82	1.57
SCS - population	1.10	1.20
SCS - ecological relationships	0.95	1.47
SCS - total	2.85	3.03
MUSEA - total	10.21	4.66

N = 379

Low correlations were found between similar themes on different scales. These are shown on this page below, as they were presented by the investigator. A conclusion made in this study was that the low correlations were evidence of an "index of the unobtrusiveness" of the MUSEA.

MUSEA Subscale Correlation Matrix

	Word Association Scale Total	Free Choice Scale	Sentence Completion Scale Total
Word association scale - total	1.00	--	--
Free choice scale - total	0.23	1.00	--
Sentence completion scale - total	0.21	0.17	1.00

Significant at the 0.01 level of confidence

N = 379

Product moment correlations were calculated for each theme between scales of the MUSEA and are shown on the following page.

MUSEA Theme Correlation Matrix

	SCS Pollution	SCS Population	SCS Ecological Relationships	SCS Total	FCS Total	WAS Total
WAS-pollution	0.37					
WAS-population		0.72				
WAS-ecological relationships			0.15			
MUSEA-total	0.38	0.44	0.45	0.80	0.60	0.62

Significant at the 0.01 level of confidence

Multiple regression analysis indicated no significant difference between scores on the MUSEA from small, medium, or large communities. Likewise, no significant difference was found between scores on the MUSEA for male and female subjects.

Interpretations

The following conclusions were drawn by the investigator and were presented under the discussion section of his paper.

1. An unobtrusive environmental attitude instrument with acceptable reliability and construct validity was developed.

2. This instrument was "successful in its unobtrusiveness to the extent that, in the opinion of teachers, a majority of the 379 subjects were indeed unaware of what type of test they were taking or of its intent." The author of this paper, however, remarks that to conceal the identity of the nature of the questions and to allow a wide range of responses, questions in the MUSEA were structured to be nondirective. He concludes that by minimizing the directiveness of the questions the ability of the instrument to assess themes was reduced. Consequently, this instrument assesses overall environmental attitude rather than specific attitudes as originally intended by the researcher.

3. Some individuals were more willing or able to express their attitude on one scale than another. In some cases students defined the stimulus words instead of responding to them with feelings or emotion. The researcher recommends that all three scales of the MUSEA be used in assessing a subject's attitude.

4. The Word Association Scale was found to be the most unobtrusive subscale in this instrument. While this subscale was the most open, it was also the most difficult to score. Many responses had to be scored as neutral.

5. The Free Choice Scale appeared to assess attitudes consistent with the entire MUSEA. The advantages of the FCS are its ease of administration and scoring. While the FCS does allow considerable freedom, it offers more direction and is easier to interpret than the SCS and WAS.

6. The Sentence Completion Scale was easier to score than the WAS (it allows less divergent responses than the WAS) but the disadvantage is that an astute subject is more likely to see through the guise of the SCS than the other scales, thus revealing the nature of the MUSEA.

ABSTRACTOR'S ANALYSIS

Attitude measurement can occur within several dimensions. Assessments can be made of subjects' perceptions or of specific behaviors they display. For example, a person might state "I hate smoking. It is a nasty habit and it turns me off." Or he might respond to the topic by saying "Smoking is for relaxing. I love to be around someone who smokes a pipe." These are examples of two possible perceptions one might have toward smoking and each could be expressed verbally or by paper and pencil means. On the other hand, a person's attitude toward smoking could be assessed by watching his behavior in a situation where someone offers him something to smoke. His/her behavior in this setting might indicate how he/she feels toward the act of smoking.

Stimuli which elicit feelings or emotions may be either artificial or natural. Events can be structured in which questions are asked or statements are made in order to produce specific responses. A person, for example, can be interviewed or asked to respond to a questionnaire in which case the format is artificial or obtrusive. Alternatively, a person's behavior can be recorded in a more natural or unstructured setting, in which case the stimuli become less obtrusive (that is to say, the respondent is less aware or perhaps unaware altogether that specific attitudes are being assessed).

The following diagram depicts the two dimensions described above. From this matrix it can be seen that attitudes are expressed as multidimensional, and that different measures are needed in order to record the various responses.

Indicators of Feeling

	Perceptual	Behavioral
Setting	A	B
Artificial (obtrusive)	(Attitude questionnaire assessing feelings toward smoking)	(Role-playing activity designed to probe feelings toward smoking)
Natural	C	D
(unobtrusive)	(Listening to student comments after class that suggest feelings toward smoking)	(Watching students in groups away from school where cigarettes are available)

The following scheme represents a few thoughts I have developed recently on the multidimensional nature of attitudes and how these feelings may be potentially assessed. The researcher in this study has developed and shared with readers new techniques for measuring environmental attitudes. He has demonstrated how three methods (word association, sentence completion and a free choice scale) can be used to increase the "naturalness" or "unobtrusiveness" of the format (setting) employed in which attitudes can be measured. The Moyer Unobtrusive Survey of Environmental Attitudes represents methodologies that can be used to measure attitudes in science education in ways different from most of the common paper and pencil, self-report techniques currently being used today. In this regard, the investigator has led the way into a potentially fresh, new direction of attitude research in our field. The degree to which obtrusiveness vs. unobtrusiveness influences the validity of student responses during attitude assessment is a research question of the highest order. This study exposes the question and should serve as a catalyst for further investigation.

This study is well-written and is easy to follow. The statistical methods used to analyze the data appear appropriate and are clearly communicated to the reader. Prior work with the assessment techniques used are referenced and implications are discussed. The major weakness in the writeup of this study is that the author does not include examples of items contained in the MUSEA. I found it difficult to evaluate the techniques that were being forwarded without having access to any of the items--or at least examples of the items. The construction of items for an attitude instrument is a difficult task, one that requires experience or at least considerable help from experts. In Edward's Techniques of Attitude Scale Construction (1957), for example, several conditions that should be met are delineated. Readers of this investigation have not been exposed to the processes used for item construction and selection, nor have they been given an opportunity to glimpse the content of the items.

One additional concern I have involves a set of questions that can, of course, be asked of any study--that of validity. While construct validity is claimed by the investigator, questions of content, concurrent and predictive validity are not mentioned and remain unresolved. Since these are seldom established in a single study, it is important for researchers in this area to expose these unresolved parameters and to suggest further studies that will help eliminate these deficiencies. I would recommend further work with the MUSEA before suggesting its use with various populations.

As I previously stated, this study represents a potentially fresh, new direction in attitude assessment. By comparing how students respond to different attitude instruments, we shall be able to unlock many of the secrets that baffle those of us who are interested in this area of research. This study represents an excellent attempt to learn more about how student perceptions may be influenced *vis-a-vis* different measurement techniques and settings. Though more work needs to be done perfecting the techniques forwarded in the MUSEA, this study adds another important link to the ever-growing field of attitude research in science education.

REFERENCES

Corey, S. M. "Professed Attitudes and Actual Behavior." Journal of Educational Psychology 28: 271-279, 1937.

Edwards, A. L. Techniques of Attitude Scale Construction. New York: Appleton-Century Crofts, 1957.

Fry, E. "A Readability Formula that Saves Time." Journal of Reading 513-516, 575-578, 1968.

Kerlinger, F. N. and J. P. Elazar. Multiple Regression in Behavioral Research. New York: McGraw-Hill, 1954.

Oppenheim, A. N. Questionnaire Design and Attitude Measurement. New York: Basic Books, Inc., 1966.

Sacks, J. M. and S. Levy. "The Sentence Completion Test." In L. E. Abt and L. Bellak (eds.) Projective Psychology. New York: Alfred A. Knopf, 357-402, 1950.

Crawley, Frank E. and John W. Shrum. "Effects of Learning Structure Condition on Change in Preference for Science Courses." Journal of Research in Science Teaching, 14(3): 257-262, May 1977.
Descriptors--Biological Sciences; *College Science; *Educational Research; Higher Education; *Instruction; Physical Sciences; *Science Courses; Science Education; *Student Attitudes; *Student Motivation

Expanded abstract and analysis prepared especially for I.S.E. by Richard M. Schlenker, Maine Maritime Academy.

Purpose

The purpose of this investigation was singlefold. Crawley and Shrum strove to ascertain whether there was a significant difference between students' attitudes toward the area of science studied when the learning environment provided was compatible with the perceived preferred learning environment and when the environment provided was incompatible with the perceived preferred environment. In this regard the authors hypothesized the difference not to be significant in each of four introductory science areas: biology, chemistry, geology and physics.

Rationale

Recent investigations have suggested a relationship between instructional environment and attitude toward the subject being studied. Student attitudes toward a particular discipline seemed most positive when the instructional milieu matched perceived preferred learning environments and/or life style orientation. Environments not matching students' preferences were incongruous with positive attitudes.

Once either positive or negative course attitudes had been developed future confrontations with the specific subject tended to elicit relatdd overt actions. When confronted with having to enroll in one science course or another, students demonstrate strong preferences for those areas having formerly stimulated a positive orientation.

Conversely, students possessing the latter attitude would not be expected to show preference for courses they perceived requiring them to learn in ways inconsistent with perceived preferred styles.

A specific model was not applied. However, the inference is made that the findings reported in the literature are applicable conceptually across a broad range of subject areas.

Research Design and Procedures

One hundred fifty-three students enrolled in introductory courses in biology (56), chemistry (26), geology (48), and physics (23) at the University of Georgia were administered the Structural Compatibility Inventory (SCI) just prior to the final examination and the Subject Preference Scale (SPS) at the beginning of the course and just prior to the final examination during the spring quarter of 1975. The SCI measured the extent to which students were learning in preferred ways while the SPS indicated the degree of preference students had for the science course they studied. Additionally, pre-posttest administration of the SPS provided subject preference gain scores.

Two course types were used in each science area; (1) a course established for elementary education majors; (2) a regular introductory course available to students having varying degree and career interests. Accordingly, the SCI and SPS were administered to students enrolled in eight separate science courses.

Students were lumped into biology, chemistry, geology and physics categories based upon course enrollment and subsequently segregated into eight subject matter specific compatible and incompatible subgroups based upon the SCI results. An independent groups two-tailed t-test was then used to compare mean gain scores (derived from SPS pre-posttest results) between subgroups in each subject area.

Findings

The original hypothesis was rejected for the physics subject area. There was a significant difference ($p < 0.01$) in gain scores between students in preferred and non-preferred learning environments; the preferred learning environment subgroup showed the largest mean gain score. Subject matter preference gain scores did not differ significantly between subgroups in the biology, geology and chemistry areas.

Interpretations

The following conclusions were drawn based upon the results of the study:

1. The evidence at hand supports the contention that positive attitudes toward science content are associated with those who learn in preferred ways.

2. The cognitive dissonance theory as a possible rationale for science course preference (it would be inconsistent for students to demonstrate strong preferences for courses in which they were expected to learn in ways not preferred) is supported.

3. Different learning environments should be provided between different sections of science courses offered at the introductory level. Where enrollments at the introductory level are small, instructors should provide a variety of learning environs within the same course.

ABSTRACTOR'S ANALYSIS

During the past few years there has been increasing interest in affective parameters of education (Renner, et al., 1978). According to Renner, et al. (1978) those affective parameters continuing to influence educational research are attitudes, beliefs, self-concepts, values, and interests. The study discussed herein is a small but important segment of the affective domain's attitude subconstellation.

Attempts to provide a multiplicity of learning environments in a single educational setting or to accommodate the several learning styles existent within a single heterogeneous group of students have long been a goal of educators. In the early 70's and before, the primary interest was in evaluating environments thought to be the best facilitators of learning gain across wide varieties of backgrounds, levels of reasoning ability and the like (Postlethwait et al., 1977). However, research concerning attitudes toward subject matter as a function of learning environment was seldom broached. Yet, one of science education's primary goals in recent years has been to foster positive student attitudes toward science!

This study's uniqueness stems from the subtle underlying assumptions of its authors. First, students leaving introductory science courses with positive attitudes toward the course will be likely to delve more deeply in that science area at some future date whereas the opposite might be expected of those departing with negative attitudes. While this contention was not evaluated in the study, it was shown that matching actual and perceived preferred learning environments did positively affect students' attitudes about physics. The logical hypothesis therefore is students (from this study) who were in a perceived preferred learning environment might be likely to take additional physics courses when the opportunity unveiled itself.

Several contemporary studies have dealt with preservice and inservice teacher attitudes toward teaching methodologies as well as science itself (Jaus, 1978; Gabel and Rubba, 1979; Piper and Hough, 1979; Bratt and DeVitro, 1978; Lazarowitz, Barufaldi and Huntsberger, 1978). Since Crawley and Shrum included science courses intended for elementary education majors, their study adds another dimension to the work already accomplished. The implicit suggestion that positive attitudes on the part of preservice elementary teachers about science produces teachers interested in teaching science is that added dimension. Further, teachers of this nature probably realize the value of making available a variety of learning environments for any given class of students.

Finally, this research suggests the avenue to education of a scientifically literate society. Positive attitudes on the part of today's students regardless of their career goals may lead tomorrow to a society willing and able to evaluate scientific issues from a position of understanding.

For these reasons Crawley's and Shrum's research is one of the most important of the present attitude research matrix.

The degree research results are generalizable depends upon the number of variables controlled while the work is conducted. When a control group is not used the way is open to argue that a study's results might have been different if one were used, even though they might not have been. The same argument exists concerning the use of random sampling techniques over the lack of using such techniques. The willingness of others to apply research results may depend upon the amount these and other variables are controlled. In short, the tighter the controls, this study notwithstanding, the greater the contribution to the field of endeavor.

Generalizability also depends upon the use of a sufficiently large sample and proper description of the sample. Failure to control these parameters often leads to only one conclusion. The results of the study are applicable only to the sample used in the study. While this study has opened a significant line of science education research, the application of its results should be used with caution until the questions of sample size and description are resolved.

There are several possible approaches to additional research in this area.

1. Studies of this type should be conducted using random sampling techniques.

2. The study should be duplicated controlling for the science background of the subjects prior to the study.

3. Studies should be conducted to include all science disciplines.

4. Studies of the same type should be conducted at other post-secondary institutions.

5. Longitudinal studies should be designed to follow students once they complete their study at the introductory science level. These studies should:

   A. Look at what students who exhibit positive attitude gains at the introductory levels do when given the opportunity to study again in the area which produced positive attitude gains.

   B. Look at effects of positive attitude gains toward science resulting from studying in perceived preferred learning environments upon teaching styles of inservice teachers. The attitudes of students having these teachers also should be examined.

REFERENCES

Bratt, M. and A. DeVito. "The Development and Field-Testing of an Inventory of Humanistic Attitudes Towards Science Teaching." Journal of Research in Science Teaching, 15(6): 551-557, 1978.

Crawley, F. E. and J. W. Shrum. "Effect of Learning Structure Condition on Change in Preference for Science Courses." Journal of Research in Science Teaching, 14(3): 257-262, 1977.

Gabel, D. and P. Rubba. "Attitude Changes of Elementary Teachers According to the Curriculum Studied During Workshop Participation and Their Role as Model Science Teachers." Journal of Research in Science Teaching, 16(1): 19-24, 1979.

Jaus, H. H. "The Effect of Environmental Education Instruction on Teachers' Attitudes Toward Teaching Environmental Education." Science Education, 62(1): 79-84, 1978.

Lazarowitz, R.; P. J. Barufaldi; and J. P. Huntsberger. "Student Teachers' Characteristics and Favorable Attitudes Toward Inquiry." Journal of Research in Science Teaching, 15(6): 559-566, 1978.

Postlethwait, S. N., et al. Exploring Teaching Alternatives. Minneapolis, Minnesota: Burgess Publishing Company, 1977.

Renner, J. W.; M. R. Abraham; and D. G. Stafford. A Summary of Research in Science Education-1976. New York: John Wiley & Sons, 1976.

Fraser, Barry J. "Selection and Validation of Attitude Scales for Curriculum Evaluation." Science Education, 61(3): 317-329, 1977.

Descriptors--Affective Objectives; *Attitudes; *Curriculum Evaluation; Educational Research; *Evaluation Criteria; *Science Education; Tests; Test Selection; *Test Validity

Expanded abstract and analysis prepared especially for I.S.E. by John E. Penick, University of Iowa.

Purpose

This article describes criteria to aid educational evaluators in the selection, modification, and validation of scales for curriculum evaluation. In illustrating these processes the author provides a rather full description of five attitude scales which he sees as potentially useful but not well-known.

Rationale

Working primarily with scales in the affective domain, the author expressed a great deal of concern for the proliferation of new scales. He felt that evaluators should judiciously select scales on set criteria rather than develop new instrumentation.

In selecting a battery of scales for curriculum evaluation, consideration should be made of: educational importance, multidimensionality, and economy. Educational importance (Cronbach, 1971) "demands that each aim measured in a battery of scales be individually educationally worthwhile and that, also, the battery as a whole neglect no relevant aim of major technical import." Economy is a measure of length, with lengthy scales being generally considered inappropriate if a battery of scales is being contemplated.

Research Design and Procedure

A review of the literature in science education led to the identification of 117 articles which stated 1,547 aims considered desirable for science education. Two hundred and seventy-six of these aims were attitudinal. Each was classified according to Klopfer's (1971) categories. Categories one through five were seen as being of sufficient educational importance to include in the contemplated battery items measuring aims in each of these categories.

Table 1

Five Affective Aim Categories in Klopfer's Classification, and Percentage of Affective Aims Stated in the Literature Falling into Each Category

Category	Title	Percentage of Stated Aims
H.1	Manifestation of favourable attitudes toward science and scientists	16
H.2	Acceptance of scientific inquiry as a way of thought	12
H.3	Adoption of "scientific attitudes"	32
H.4	Enjoyment of science learning experiences	17
H.5	Development of interests in science and science-related activities	18
H.6	Development of interest in pursuing a career in science	5

Category six was considered to be of such lesser importance as to not be critical for the battery. From this, it was determined that the three criteria of educational importance, multidimensionality, and economy would be best met using five relatively short attitude scales, with each scale measuring one of Klopfer's aim categories.

Five attitude scales were chosen to adequately cover the various dimensions, consider each of the five aims, and be of short length. All five of the scales were developed in Australia or England and showed varying reliabilities on different formulations ranging from 0.53 to 0.90.

The first scale in Table 2 is a modified version of a scale developed by Ormerod (1971) to measure attitudes toward the social implications of science, an especially important aspect of the evaluation of the contemporary science curricula.

Table 2

Five Attitude Scales, Together With the Klopfer Category of, the Number of Items in, and Reliability of Each Scale

Attitude Scale	Klopfer Category	No. of Items	Cronbach Reliability Validation (N = 165)	Cronbach Reliability Cross-Validation (N = 1,158)
Social implications of science	H.1	8	0.81	0.77
Attitude toward inquiry	H.2	8	0.67	0.72
Adoption of scientific attitudes	H.3	11	0.63	0.50
Enjoyment of science lessons	H.4	7	0.85	0.81
Interest in science outside lessons	H.5	6	0.80	0.79

The second scale in Table 2, based on a sub-scale of Meyer's (1969) "A Test of Interests," is one of the few existing instruments designed to measure Klopfer's category H.2.

The third scale of Table 2 is a modified version of TOPOSS (Test of Perception of Scientists and Self), developed by White and MacKay (1976) and measuring pupils' adoption of attitudes like curiosity, suspended judgment, etc. The last two scales of Table 2, measuring enjoyment of science lessons and interests in science outside lessons, respectively, were adapted from scales developed by the Schools Council Project for Evaluation of Science Teaching Methods (1973) from original scales developed by Laughton and Wilkinson (1965).

Each original item in the five scales was checked for face validity and the presence of ambiguities by a panel of people with expertise in measurement and science education. Reading levels were considered and some items were rewritten or deleted.

After modification, the battery of scales was administered to an Australian seventh grade sample. These 165 students in six schools provided data for statistical indices for identifying faulty items and describing the validity of the refined scales after removal of faulty items. Cross-validation of the scales involved giving the battery to 1,158 seventh grade pupils in 46 high schools in Australia. Internal consistency was measured with a positive item-remainder correlation significantly different from zero at the .05 level. Each item failing this criterion of internal consistency was removed. Cronbach Alpha reliability coefficients of internal consistency ranged from 0.63 to 0.85 with a median of 0.80 for the validation study and from 0.50 to 0.81 with a median of 0.77 for the cross-validation study.

Discriminate validity, an indication that each scale measures a unique construct not measured by the other scales, was also tested for by intercorrelation between scales. Scale intercorrelations were considered acceptable if they were less than the geometric mean of corresponding scale reliabilities.

Sensitivity, an index of the test's ability to detect pupil changes of the order of magnitude which actually occur was determined adequate if pupil scores covered a large range of the available score range. The present battery of attitude scales possessed such a range and was therefore considered to possess satisfactory sensitivity.

Since the ultimate usefulness of any scale is determined by correlations existing between those scales and other variables deemed important, these scales were correlated with four other variables: an instructional treatment variable, socio-economic status, I.Q. and sex. The instructional variable in tested classrooms was either use of Australian Science Education Project (ASEP) materials or alternative materials having been used in science classes in the eight months prior to administration of the scales. Socio-economic status was determined with Congalton's occupational classification and I.Q. was measured with a version of the Otis test. Statistical analysis was performed on 343 sub-groups of individuals rather than the 1,158 pupils in the whole sample.

Findings

Pupils who had used ASEP materials were found to express more favorable attitudes toward science than pupils who had used non-ASEP materials on both the social implications of science scale and the enjoyment of science lessons scale. High socio-economic status pupils were also found to have more favorable attitudes than did lower SES pupils on the social implications of science scale and the adoption of scientific attitude scale. I.Q. was significantly positively related to poorer performance on attitude toward inquiry and adoption of scientific attitude scales, a finding consistent with prior results. On three attitude scales, adoption of scientific attitudes, enjoyment of science lessons, and interest in science outside lessons, boys tended to exhibit more favorable attitudes toward science than did girls. This finding was also consistent with prior evidence. No significant correlation was found between sex and attitude toward the social implications of science.

Interpretations

This paper was designed to provide criteria to guide the selection, modification, and validation of scales for curriculum evaluation and to illustrate the application of these specific criteria to a selected battery of five attitude scales. Through this procedure, the author hoped to make these five English and Australian attitude scales better known while precisely illustrating the points he wished to make in the paper. After identifying the important characteristics of attitude scales, the author proceeded to demonstrate how specific scales could be shown to meet the various criteria. In doing so, several modifications were made to the original scales. These revised scales were then tested for internal consistency, discriminate validity, and sensitivity during both a validation study and a cross-validation study. In addition, correlations between scores on each scale and an instructional variable, socio-economic status, general ability, and sex, were calculated.

ABSTRACTOR'S ANALYSIS

A large percentage of recent publications include some measure or reporting of attitudes. In many instances, the author develops a new attitude measure because existing measures are not considered adequately sensitive to the interests of that researcher, selects a panel of experts to provide validity of the instrument, and proceeds to administer the instrument to the selected sample population. Aside from the obvious difficulty of completely developing a new, valid, reliable, and sensitive attitude scale, the ever expanding pool of attitude scales is beginning to make generalizations and comparisons between studies virtually impossible. Fraser's suggestion that they could more profitably be concerned with the judicious selection of existing scales, modification of chosen scales to enhance suitability for use in a particular study and the validation of modified scales, is quite accurate.

Fraser's concept and technique of selecting and evaluating attitude scales for use in science education is quite useful and workable. He clearly demonstrates that the quality of a scale is directly related to the criteria used in selecting that scale. Further, he has a clear idea of the various criteria which are critical to scale selection, use, and development.

The three criteria (educational importance, multidimensionality, and economy) recommended for use in an initial selection of an attitude instrument were defended. After selecting scales to fit each of Klopfer's categories deemed important by the education aims survey, Fraser proceeded to determine the standard statistical criteria of internal consistency, discriminate validity, and sensitivity, for all five modified scales.

In this article, Fraser provides more than a clear, concise rationale and technique for selecting, developing, and evaluating attitude scales in science education. He has also clearly demonstrated the application of this technique while providing data on five previously unknown scales which may ultimately prove useful in science education research.

REFERENCES

Cronbach, L. J. "Test Validation," in R. L. Thorndike (Ed.) Educational Measurement (2nd edition). Washington: American Council on Education, 1971.

Klopfer, L. E. "Evaluation of Learning in Science," in B. S. Bloom; J. T. Hastings; and G. F. Madaus (Eds.) Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Laughton, W. H. and W. J. Wilkinson. "Pupils' Attitudes to Science Teaching." Education in Science, 26: 31,33, 1968.

Meyer, G. R. A Test of Interests. Milton, Queensland: Jacaranda Press, 1969.

Ormerod, M. B. "The 'Social Implications' Factor in Attitudes to Science." British Journal of Educational Psychology, 41: 335-338, 1971.

Schools Council Project for Evaluation of Science Teaching Methods. "A Modified Version of the Science Pupil Opinion Poll." University of Leicester, 1973.

White, R. T. and L. D. MacKay. Tests of Perception of Scientists and Self. Melbourne: Australian Council for Educational Research, 1976.

Novick, S. Shimshon and D. Duvdvani. "The Relationship Between School and Student Variables and the Attitudes Toward Science of Tenth-Grade Students in Israel." Journal of Research in Science Teaching, 13(3): 259-265, 1976.

Descriptors--*Educational Research; School Role; Science Education; Secondary Education; *Secondary School Science; *Scientific Attitudes; *Student Characteristics

Expanded abstract and analysis prepared especially for I.S.E. by Michael Szabo, The Pennsylvania State University.

Purpose

The authors' purpose is to investigate and further the understanding of general scientific attitude among students in high schools of Israel. Attitude was studied as a function of type of school, academic specialization, and curriculum, as well as sex, achievement level, and cultural background of students.

Rationale

The rationale given for conducting this descriptive study was loosely based upon an inferred relationship between attitude and "educational frameworks," differentiated science achievement, and new science as curricula adapted to Israel. Conflicting research was cited to support this rationale. No coherent or recognized theory or model of attitude formation was explicated. The major implication appeared to be that exposure to science at the high school level will directly impact the development of scientific attitude.

Research Design and Procedures

The design involved random selection of strata of 25 high schools (684 tenth-grade Ss). It was basically a one-group, posttest-only design and provided a different analysis of the data collected and reported in Novick and Duvdvani (1976).

Data were collected from school records and the Scientific Attitude Inventory (SAI). This instrument was designed to assess intellectual and emotional components of scientific attitude. Although the design lends itself to data analysis via ANOVA, and "significant main and interaction effects" were noted, no indication was given that any statistical tests were performed.

Findings

The findings are complex as various combinations of the six stratifying variables were mixed in an incompletely crossed design. For example, the first set of results reported on school type as the major variable with sex and cultural background as "secondary variables." The effect was assessed on both emotional and intellectual attitude. Sex had no effect on attitude but school type and culture did. An interaction between sex and school type was reported relative to intellectual attitude.

Interpretations

The authors conclude that five of the six variables affect student attitudes. Further, (1) religious schools do not change science's image, (2) agricultural students are less positive emotionally toward science, (3) students of Western extraction hold more positive attitude than those of Eastern extraction, (4) future science majors and high achievers are more positive, and (5) exposure to new curricula does not improve science attitudes.

ABSTRACTOR'S ANALYSIS

The results and conclusion must be tempered with certain methodological and conceptual considerations.

Internal Validity. The SAI yields a total score which appears to be relatively sound with substantial reliability and a reasonable construct validity. However, although the authors of the instrument used the two separate subscales to assess emotional and intellectual attitude toward science, no evidence is presented that these constructs indeed exist or that the items that measure them have content validity. No reliability estimates of the two scales were presented or referenced anywhere. This problem places severe limitations on the conclusions. Other limitations relative to the SAI are noted by Szabo (1979).

A further limitation is that the reliability and validity data were obtained prior to the translation into Hebrew. The introduction of cultural biases and failure to reestablish reliability and validity in the new setting is seen by the reviewer as a severe limitation.

Many rules of scientific reporting are broken in this report. Although significant main and interaction effects are reported, there is no description of the statistical tests and less than minimal data appear. Only selected means and no standard deviations are presented. For example, there is no way for the reviewer to check the statement that boys assume more positive emotional attitude ($\bar{X} = 61.9$) than girls ($\bar{X} = 60.3$). Without standard deviations and sample sizes, the significance of a mean difference of 1.6 cannot be checked.

If the sampling unit was either the school or the intact classrooms, the appropriate statistic to use would be school or class mean, rather than individual students' scores. Such unit sampling requires different computations than for individuals as the sampling unit (Walker and Lev, 1953). This remains an unknown quantity in this study.

The method of subdivision of the sample has questionable reliability. Achievement level was defined as percentage (not an equal interval scale measure) of final grade (notoriously unreliable measures). Classification as to science curriculum assumes clear distinction between the newer sciences and the "traditional" sciences. Such differences have been elusive to validate in these curricular categories in the United States.

The authors'implications are not supported by the data analysis and should not be associated with this research study. For example, the implication that intellectual attitude needs "explicit educational attention" is simply not supported by these data. That intellectual attitude is even measured by SAI is open to debate. The second implication calls for more relevant (to students' interests) science teaching for humanities majors. Since humanities majors were not specifically addressed in this study, the reviewer questions what authority supports this conclusion.

An Alternative Plan. How could a study be designed to yield meaningful results that would contribute to our knowledge of attitude constructs and their interplay with the educational framework? The reviewer would like to make several suggestions.

First, an analysis of various theories or models of attitude formation must be conducted. A great deal is known, for example, about the persuasive communication (Shrigley, 1978) and the cognitive dissonance (Festinger, 1957) models of attitude. The former is used in science teaching (often unknowingly) and has been abstracted to science teaching by Shrigley.

Next, the structure of the educational framework must be dissected to determine the extent to which it contains components in sympathy with the components of the model of attitude used. For example, the communication persuasion model clearly shows that the credible source communicator should present both sides of an issue to intelligent audiences to foster attitude change (Aronson, 1976). The unit of PSSC physics which deals with the model of light does a credible job in presenting the particle and wave models, as well as leading the student to formulate his/her own conclusions.

The result should be the emergence of a logical rationale which suggests why or why not an educational framework can be expected to contribute to attitude formation.

Third, individual variables that are related to attitude scores must be reliable, and validity controlled or measured in the analysis. For example, females seem to be more susceptible to persuasive communication relative to attitude change than are males (Cohen, 1964). And the relationship between academic ability and attitude (in a correlation rather than a causative sense) is well established.

Fourth, attitude instruments should be designed which can be related to the components of the model, the educational framework, and individual learner differences. The validity of the instrument must be established relative to both the constructs and the content. Then and only then can we be hopeful of meaningful results which are amenable to interpretation.

The above plan cannot be completed as a doctoral dissertation. It will take a long-term effort by an individual or concentrated efforts by a dedicated team. The result should be, however, a deeper understanding of science attitude formation and a knowledge of how to build educational structures to foster scientific attitude without inducing unwanted side effects (e.g., severe decrements in knowledge and process acquisition.

REFERENCES

Aronson, E. The Social Animal. San Francisco: W. H. Freeman, 1976.

Cohen, A. Attitude Change and Social Influence. New York: Basic Books, 1964.

Festinger, L. A Theory of Cognitive Dissonance. Stanford, Calif.: Stanford University Press, 1957.

Novick, S. and D. Duvdvani. "The Scientific Attitudes of Tenth-Grade Students in Israel, as Measured by the Scientific Attitude Inventory." School Science and Mathematics, 76: 9-14, 1976.

Shrigley, R. L. "The Persuasive Communication Model: A Theoretical Approach for Attitude Change in Science Education." Journal of Research in Science Teaching, 15: 335-341, 1978.

Szabo, M. Abstract and Analysis of "The Scientific Attitudes of Tenth-Grade Students in Israel, as Measured by the Scientific Attitude Inventory." Investigations in Science Education, 5(2): 25-30, 1979.

Walker, H. M. and J. Lev. Statistical Inference. New York: Holt, Rinehart and Winston, 1953.

Kauchak, D. P. "The Effects of Essay Writing on the Attitudes of Undergraduate Science Methods Students." Journal of Research in Science Teaching, 14(2): 139-143, 1977.

Descriptors--*Affective Behavior; Attitudes; Changing Attitudes; College Students; *Educational Research; *Essays; Higher Education; *Instruction; Science Education; *Teacher Education

Expanded abstract and analysis prepared especially for I.S.E. by Michael Szabo, The Pennsylvania State University.

Purpose

Kauchak (1977) reports two experimental studies which address the question of attitude development in preservice science teachers at the elementary and secondary levels. In the first experiment, the hypothesis tested was that writing an essay favorable toward a topic increases one's attitude toward that topic. In the second experiment, the hypothesis of concern was that differing amounts of reward for an essay affect attitude toward the essay topic.

Rationale

The rationale for Experiment I was based upon research done in the 1950's relating essay writing with changed attitude. The second experiment was predicated on research from the middle 1960's which suggests a positive relation between reward and attitude change.

Research Design and Procedures

Experiment I

The design featured random assignment of Ss to one of two treatment conditions (essay writing vs. nonessay writing) or control. Only the two treatment groups were posttested on attitude toward the essay topic (Bloom's Taxonomy); only the control group was pretested for baseline data.

The independent variable was an essay writing examination technique to increase attitude while the dependent variable consisted of scores on a 20-item Likert-type instrument designed to measure attitude toward Bloom's Taxonomy. Construct validity, established through a panel of judges, and reliability data were reported.

The Ss were 112 undergraduate secondary methods students enrolled in a methods course. No breakdown by sex, age, or other variable was provided by the author.

The experimental procedure involved having the treatment group write an essay in favor of Bloom's Taxonomy as part of the examination for a self-instructional module on that topic. The nonessay group classified and wrote objectives using the Taxonomy.

Experiment II

The design involved random assignment of Ss to one of four groups (1) control, (2) essay in favor of topic for four points, (3) essay in favor of topic for two points, and (4) essay on disadvantages of the topic for two points. The total point value of the test was 30. It was hypothesized that writing an essay would alter attitude toward the topic in direct proportion to the amount of reward (Group 4 > Group 1 > Group 3 > Group 2), where scores are inversely related to attitudes. The independent variable, amount of reward in conjunction with essay writing, had three levels.

The dependent variable was the score on a 10-item Likert-type instrument measuring attitude toward the topic, in this case inquiry mode of teaching. The test reliability was reported but no validity information was presented.

The Ss were 106 elementary undergraduate methods students.

Findings

Using t-tests, it was found for Experiment I that the essay writing group scored lower on the attitude test (low scores imply high attitude and vice versa) than the control group and the nonessay group. Furthermore, the nonessay group had lower attitude scores than did the control group. Kauchak concluded that writing an essay favorable to a topic in a test situation increases attitude toward that topic.

The findings, based upon pairwise t-tests, indicated that for Experiment II the group which wrote the favorable essay for two points (Group 3) had a more positive attitude than students in either the control group (Group 1) or the group writing on disadvantages (Group 4) for two points. No other differences were significant.

Interpretations

Kauchak concluded that writing an essay in a test situation can change attitude in the direction of the position advocated by the essay.

ABSTRACTOR'S ANALYSIS

The researcher has done a creditable job on many counts, two of which the reviewer will highlight. The topic of attitude development is quite timely as it recognizes the need to develop attitude toward science in citizens through public education. In addition, the experimental nature of these studies permits causal inferences between the independent variable of essay writing and the criterion of attitude.

Rationale. The reviewer would like to comment on the rationale of the studies to clarify issues for future research.

The studies do not seem to be couched in any theory or model of attitude development. A theoretical base is seen as a necessary condition for precise hypotheses, valid treatments, and insightful interpretation.

The reviewer infers that the persuasive communication model of attitude development (Hovland, et al., 1953; Shrigley, 1978) applies to Kauchak's research. This model argues that, when presented with formal communication containing pertinent information (implying a need for attitude change), rational humans will acquire different dimensions of attitude.

Based on parts of this model, the treatments seem well designed. Based on other parts, however, they seem counterproductive. Zimbardo and Ebbesen (1969) have shown that the recipients in an intelligent audience should not have conclusions drawn for them. Rather, attitude change is more likely when they are allowed to draw conclusions themselves. This appears consistent with the essay writing treatment. Kauchak also used a credible source (the instructor) which is more effective in bringing about attitude change (Cohen, 1964).

The second experiment did not conclusively support a relationship between rewards and attitude formation. Zimbardo and Ebbesen (1969) showed that rewards causing people to respond to a persuasive communication may be direct or anticipated. By focusing on the points awarded for essays, Kauchak ignores other perhaps more subtle rewards. For example, responding in a manner consistent with the perceived values of the credible source would be rewarded with a better score. This anticipated reward argument gains plausibility when one recalls that the essay was worth at most four points out of 30. The reviewer contends that if the anticipated rewards were attended to, a more valid picture of attitude and rewards would be revealed.

Internal Validity. The design could be reviewed in terms of other features of the persuasive communication model. Most of the instruction was favorable to the topic. Aronson (1976) has shown that both sides of an argument should be presented to an intelligent community for maximal attitude development. An alternative hypothesis is that attitude scores were in general elevated due to the credible source argument (Cohen, 1964) and that the nonessay treatment may have depressed attitude scores (Ss in the nonessay group were told they might write an essay exam but then did not do so). This effect might have been estimated if the pretest-only group was also posttested. An additional test could have been made in a delayed posttest design, since attitude that stems from a credible source tends to be short-lived (Kiesler, et al., 1969).

Aronson (1976) has also shown that greater attitude change occurs when the initial position of the source is discrepant from the recipient. Initial attitudes were not assessed in the study. An alternate design which would have provided these data and in addition permitted a test of the pretest--treatment interaction is the Solomon 4 Group design. The sample size apparently was sufficient to permit this design with adequate power.

Methodologically, the study could be improved in terms of statistics, instrumentation, and hypothesis clarity.

The use of multiple t-tests does not control the family-wise level of significance. Hence some of the tests may have been conducted at probability levels considerably larger than .05. A more desirable procedure involves an overall F-test followed by an appropriate a posteriori test (Winer, 1962).

The validity of the criterion test can be questioned. The use of the first instrument (Taxonomy) has not been replicated, a requirement for validity--the test of time as it were. Judging of items does not deal with construct validity unless the domain of observable behaviors of the construct (in this case, one of high inference) is specified (Nunnally, 1967). Kauchak provides no evidence that the domain was identified for either of the criterion instruments. The second instrument (Inquiry) has no reported validity at all. Although what these instruments measure may be in question, the reliability of the measurement seem sound.

The hypothesized relationship in Experiment II is inconsistent with the rationale which predicts that any essay writing is better than none regarding attitude formation. Those who wrote an essay against inquiry teaching according to the rationale should have higher attitude than the control group. Kauchak hypothesized the reverse.

External Validity. The generalizability of the findings may be limited by the reactive testing effect. That is the attitude test scores may have been influenced by the essay writing exercise immediately preceding. This question could have been answered by the Solomon 4 Group design mentioned above.

Another note regarding the external validity of the findings is in order. Females seem to be more susceptible than males to persuasive communication (Cohen, 1964). Unfortunately, the ratio of females to males is not described, limiting the generalizability of the findings. This ratio is probably weighted in favor of females for Experiment II and in favor of males in Experiment I.

The reviewer's major suggestion is to reanalyze Kauchak's findings and rationale in terms of recent work on attitude formation (Shrigley, 1978). Specifically, the theory (or alternative theories) underlying attitude formation should be studied thoroughly by future researchers if we are able to make significant strides in our research.

REFERENCES

Aronson, E. The Social Animal. San Francisco: W. H. Freeman, 1976.

Cohen, A. Attitude Change and Social Influence. New York: Basic Books, 1964.

Hovland, C. I., I. L. Janis, and H. H. Kelly. Communication and Persuasion. New Haven, Conn.: Yale University Press, 1953.

Keisler, C., B. Collins, and N. Miller. Attitude Change: A Critical Analysis of Theoretical Approaches. New York: John Wiley, 1969.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1963.

Shrigley, R. L. "The Persuasive Communication Model: A Theoretical Approach for Attitude Change in Science Education." Journal of Research in Science Teaching, 15: 335-341, 1978.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Zimbardo, P. and F. Ebbesen. Influencing Attitudes and Changing Behavior. Reading, Mass.: Addison-Wesley, 1969.

Moore, Kenneth D. and Shirley Robards. "Field Centered Preservice Elementary Science and Content Reading Methods." Piper, M. and K. D. Moore, (Eds.) Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, 1977.

Descriptors--*Attitudes; Educational Research; Elementary School Science; *Field Experience Programs; Higher Education; *Methods Courses; *Preservice Education; Science Education; Teacher Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Michael Padilla, University of Georgia.

Purpose

The purpose of this study was to compare the attitudes of preservice elementary teachers who had experienced campus centered methods courses with those who had experienced field centered methods courses.

Rationale

Numerous criticisms directed toward teacher preparation institutions and specifically toward methods courses have been recorded. The most common complaints include the charge that methods courses do not deal with the reality of children, classrooms and teaching. Little practical experience is given in the typical program and it is this shortcoming which can be hopefully remedied in a field centered course. By integrating practical school experience with relevant theoretical material, the author hypothesized that more positive attitudes would be developed in prospective teachers.

Research Design and Procedure

Sixty-seven junior and senior preservice elementary teachers were the subjects in the study. Thirteen who had experienced both a campus centered science methods course and a campus centered content reading course were the control group. Thier attitudes were surveyed following student teaching. Fifty-four who had experienced either a science methods or a content reading course (15 were doing both), both field based, were designated as the experimental groups. These students' attitudes were surveyed immediately following their course experience but before commencement of student teaching. The Campbell and Stanley design is as follows:

X_1O X_1 = Field based science methods or content reading course
- -
X_cO X_c = Traditional campus centered science methods course

The attitude scale administered was the Robards Attitude Profile (RAP) which consisted of 16 items. Each item was a statement such as "To me the content of the course was adequate," which was followed by a four-point scale ranging from strongly agree to strongly disagree. Items 1-10 were related to the general attitudes toward the science and content methods courses, while 11-16 related to their field experiences.

Data from the first 10 items were analyzed using two separate one-way analyses of variance comparing those who had had the field centered methods to the controls and those who had had the field centered content reading to the controls. The mean attitude ratings on Items 11-16 for the field centered methods and reading courses were also reported.

Findings

For the field centered science methods course, four of the first 10 items showed significantly better attitude ratings by the field centered group, while none favored the campus centered group. The four items were:

- -- To me the content of the course was adequate
- -- To me the clarity and purpose of the assignments in the course were reasonable.
- -- Overall, the course will be useful to me as a beginning teacher
- -- The instructor encouraged students to think independently

For the field centered content reading methods group, 5 of the first 10 items indicated more positive attitudes held by the field centered group. None favored the campus centered group. The five items were:

- -- To me the objectives of the course were clearly stated
- -- To me the content of the course was adequate
- -- Overall, the course will be useful to me as a beginning teacher
- -- The teaching techniques used in the course were similar to other courses I have taken at this university

The mean attitude scores on Items 11-16 which evaluated the field component of each course were highly positive for most items. The students felt the field work was appropriate to their program of study, that it enhanced their professional growth and that the field experiences should be used in future courses. Most felt that each course needed more structure, however.

Interpretations

The authors state that the results of the "study tend to support the conclusion that involvement in field experiences leads to more positive attitudes toward methods courses." They also felt that the evidence from their study indicates that a variety of different classroom situations will help preservice elementary teachers gain experience and develop needed skills.

ABSTRACTOR'S ANALYSIS

In general the authors' attempts to document change in attitude toward their two courses is a laudable one. All too often educators make radical changes in curriculum and methodology without assessing the outcome. We need more evaluations of course outcomes so that progress can be made in an orderly fashion without extreme changes in both philosophy and content of courses. With regard to this study, more philosophical discussion concerning the tradeoffs involved in switching from campus to field based instruction would have aided the reader in establishing all of the issues involved in the change. For example, while the authors make an eloquent statement regarding the need for field based courses, they totally ignore the effect that cutting (by 40 percent) the number of campus classroom hours might have on what was learned. Issues on both sides of the question must be addressed.

The evidence presented in this study appears to document a difference in attitude toward two courses, favoring a field based strategy as compared to a campus based mode. Some questions arise, however, when the procedures and evidence are looked at in detail. One basic difficulty is in the definition of precisely what was measured. Did the Robards Attitude Profile (RAP) measure "attitudes toward preservice elementary teachers," or "attitudes toward methods courses," or did it only evaluate the courses in question? The authors imply that attitudes toward all methods courses are measured even though this reviewer finds no evidence for making this generalization. Would a panel of experts agree that the RAP has validity relative to measuring attitudes? The authors make no mention of this issue.

The reader is also faced with the difficulty of deciding exactly which subjects were administered the RAP and at which times. The control group subjects appear to have taken the test only once. If so, how were they instructed to answer the questions since the instrument queries were to be answered relative to one course only and the control subjects had taken both courses. Perhaps they took the RAP twice.

Some of the experimental group were taking only one of the courses, others took both. Yet the numbers do not add up. The authors state that there were 54 experimental subjects, yet only 49 sets of responses are recorded for Items 1-10. Too, how were dually enrolled subjects treated (the authors state that there were 15 of these)? A more precise description of the groups would have helped the reader in deciding whether the groups were truly comparable.

Additional questions arise when the authors interpret results that were "not statistically significant." The statistical proceudres are preformed in order that chance can be ruled out as a reason for differences between groups. Speaking about differences that are not significant is therefore speaking about differences that probably do not exist within the data.

In the summary and conclusion section, the authors conclude that the evidence from the study indicates that preservice teachers should be involved in a variety of different classroom situations. This reviewer sees no evidence that could lead to that conclusion. Certainly the authors may have made other observations that indicated this to be true, but that evidence was not cited in this study. Thus, this conclusion should not be stated.

This research report does not provide many useful generalizations to its readers. Too many unanswered questions regarding the nature of the dependent measure, the experimental sample and the test administration procedures cloud the results. A more precisely written report could have clarified at least some of these important issues.

DeBruin, Jerome. "The Effect of a Field-Based Elementary Science Teacher Education Program on Undergraduates' Attitudes Toward Science and Science Teaching," in Piper, M. and K. Moore, (Eds.). Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, Ohio State University, 1977.

Descriptors--*Attitudes; Educational Research; *Elementary School Science; *Field Experience Programs; Higher Education; *Preservice Education; Science Education; Teacher Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. By David P. Butts, University of Georgia.

Purpose

Many programs for the professional development of preservice teachers include field-based components prior to the culminating student teaching experience. These field-based components vary in how early in the preservice program they are scheduled, how long a time they include and the nature of their involvement of the preservice teacher in classroom instructional tasks. Because these experiences permit the preservice teachers to be directly involved in solving real instructional problems and in solving these problems experience success, it was hypothesized that their attitude toward science and toward science teaching would be changed.

Rationale

Based on an assumption that when becoming introduced to a profession, the nature of one's experience has a substantial impact on how one feels or what one believes about that profession, the use of direct involvement in field experience should provide for positive growth in how the preservice teacher feels about science and science teaching. Assuming that one fears the unknown, the converse would be true. Lack of direct experience results in negative attitudes which themselves are based on the unknowns.

Research Design and Procedure

A pre-post test design was used with nonrandom selection of 132 college preservice teachers in their intact classes for three quarters. The Moore Attitude Scale was administered at the beginning and at the end of each quarter. (No documentation was given for the validity or reliability of the instrument.) Students were involved in planning science instruction for five weeks which was then followed by a four-week in-school implementation phase. Analysis was then made on the pre-post differences of the attitude measure using a t-test on 17 unspecified variables.

Findings

There were more positive attitudes toward science and science teaching after the field experience than before. In each of the three quarters or intact classes, this change was greater in attitudes toward science teaching than in the preservice teachers' attitudes toward science.

Interpretation

When preservice teachers are directly involved in solving problems and when they have success in solving these problems, the combined impact is to help mutual trust, respect and communication to occur between the preservice teacher and the experienced classroom teacher. This trust, respect and communication leads to positive attitudes and professional growth in the preservice teacher.

ABSTRACTOR'S ANALYSIS

In posing the question about how field-based experience can help the preservice teacher's attitude toward science and science teaching, the author has specified a significant problem for study. Assuming that field-based experiences are the solution to negative attitudes can be unwarranted. If negative attitudes persist in spite of expensive field experience, the solution should be questioned. In the introduction, however, the author leads us to expect a study in which some preservice teachers will be involved in solving real problems of science instruction in the context of the classroom and others are to be involved in solving similar problems via simulated circumstances and still others are not involved in solving problems at all. Thus a comparison of those who succeed in solving the problems will be made with those who did not succeed in terms of the dependent variable, attitudes, so that conclusions can be stated about the effectiveness of field-based experience. Such a study would also lead the reader to expect to have an operational definition of the independent variable--"problem solving" and documentation of how this variable was systematically present or absent in the experiment.

Unfortunately the design and the procedure which is used answers a quite different question. Is there a correlation between the attitudes of preservice teachers before and after a course that involves them in "planning instruction" and "implementing the instruction in a field setting"? The author does find that there is a change in attitude. Due to the design of the study, that change in attitude cannot be directly attributed to any set of variables in the treatment. The author does briefly mention three possible variables: "Space utilization," "instructional time," and "variation in group size and ability." The author does allude to "17 variables" which are unspecified or defined. Thus the conclusion that attitudes do change during a quarter is the single outcome of this exploratory study.

The question first raised by the author remains both significant and unanswered. To determine the effect field-based experience has on attitude will require a study to be done in which the specific dimension or independent variables of field-based experience are defined, systematically varied in the design and comparisons then made of the dependent variables. As science educators, our practice should reflect an empirically documented research base. This study illustrates a significant question and an intuitive first step. Teacher attitudes toward science and science teaching can change--but what can be done to facilitate that change?

Jaus, Harold. "An Analysis of the Relationship of Preservice Elementary Teachers' Attitudes Toward Teaching Science and Their Science Teaching Planning Practices," in Piper, M. and K. Moore (Eds.), Attitudes Toward Science: Investigations. Columbus, OH: SMEAC Information Reference Center, Ohio State University, 1977.

Descriptors--*Attitudes; Educational Research; Elementary School Science; Preservice Education; *Process Education; Science Education; *Science Instruction; Teacher Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by David P. Butts, University of Georgia.

Purpose

Science as a process skill is a current widespread emphasis in the curriculum options available for the elementary teacher. Are teachers' plans for teaching science influenced by their attitude toward science as itself a process endeavor?

Rationale

With the widespread availability of instructional materials emphasizing the process dimension of science, the teachers' attitude toward science is seen as a key variable in their willingness to plan process-oriented objectives and activities for their students. Much documentation is noted that elementary school science instructional programs include process-oriented outcomes. It is assumed that how these materials are used in the classroom is a function of the teacher. The teacher plans the instruction and it is assumed that planning is related to the teacher attitude toward process-oriented science.

Research Design and Procedure

A post-test only no control group design with a nonrandomly selected intact class of 60 preservice teachers was used in this study. After self-paced instruction with the integrated process skills, they were tested for personal preformance of process skills plus attitude toward these skills. In a separate task they selected 10 objectives from a collection of 10 science content and 10 science process objectives. A lesson plan for each subject was evaluated for its inclusion of science process objectives and science process activities.

Findings

Preservice teachers who scored high on the attitude scale also selected a greater number of science process objectives, wrote more science process objectives into their plans as well as included more science process skill learning activities.

Interpretation

Teachers with a more positive attitude toward science process skills used them more in their planning activities for science teaching. This presents a hopeful omen that they will be more likely to use these skills in their teaching.

ABSTRACTOR'S ANALYSIS

As described in "the Educational Encounter" (Butts, 1970), there is strong logical evidence that what a teacher does in the classroom both influences what students do and what they achieve. What teachers do is also logically linked to what they know and how they feel about the importance of using their knowledge in teaching. The author has selected one piece of this linkage--does what preservice teachers know and their feelings about that knowledge correlate with their plans for science teaching? In this nonexperimental exploratory study there is some evidence to suggest that the teacher's attitude may indeed be a significant variable.

This conclusion must be cautiously examined, however. While knowledge of, plus attitude toward, science as a process skill may well be interacting to produce lesson plans with science process skills, knowledge is a controlled variable. Missing is a clear description of the context for which the lessons were being planned. Would a teacher with a desired mastery of integrated process skills and a positive attitude toward them, use them in lesson plans for students for whom such skills or activities would be inappropriate? To what extent should the findings be tempered by the nature of the dependent variable tasks? With the absence of documentation of the reliability and validity of their measure, the process skills, or the validity documentation of the attitude measure, the reader must question how much weight to place on the conclusions that are themselves based on uncertain measures.

The introduction of the study leads the reader to focus on science learning outcomes of students. Unstated is the assumption that student learning outcomes are influenced by student activities and these activities are influenced by teacher activities which are based on appropriate plans. To ascertain if these plans correlate with a teacher's attitude is the main purpose and outcome of this study. Relating student performance to teacher variables is a significant challenge of research. This study is one of those exploratory studies that convinces us that we need now to move ahead to more experimental studies that indeed show that the teacher variables do cause student growth and understanding in science.

REFERENCE

Butts, David P., Editor. Research and Curriculum Development in Science Education, No. 2, Curriculum Implementation in Elementary School Science. Austin: Science Education Center, The University of Texas at Austin, December 1, 1970, p. 9

CRITIQUES
and
RESPONSES

Berger, C. F. "Investigation of Teacher Behavior: Interaction with New Curriculum Materials." Attitudes Toward Science: Investigations. Piper, M. and K. Moore, (Eds.) Columbus, OH: SMEAC Information Reference Center, 1977.

Descriptors--Educational Research; *Elementary School Science; Inservice Teacher Education; *Science Curriculum Improvement Project, Science Curriculum; Science Education; * Science Curriculum Improvement Study; *Teacher Behavior; Teacher Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by F. Gerald Dillashaw and James R. Okey, University of Georgia.

Purpose

The purpose of this study was to assess the effect of training in SCIS materials and SCIS teaching on the actual and predicted classroom behavior of teachers.

Rationale

The advent of the new science programs during the 1960s and 1970s saw an increase in activity-centered programs. This change in emphasis pointed to a shift from a teacher-oriented class to a student-oriented class. The basis of this study is that with this shift in curriculum emphasis, different classroom behaviors on the part of teachers are required.

The study was conducted using the concept of locus of control for analyzing classroom interactions. As defined by the researcher, locus of control means the person or group who determines what happens next in the classroom. Three loci for classroom decision-making were identified. These are: (1) teacher-oriented--the teacher initiates what happens next, (2) student-teacher cooperation--decisions are shared by the teacher and students, and (3) student-oriented--students decide what happens next.

Two assumptions were made by the researcher. First, locus of control patterns differ among different science programs and, second, change in teacher behavior could be detected by a shift in the pattern of locus of control.

Research Design and Procedure

Six questions (referred to by the researcher as six studies) were posed in the investigation. For each of the studies the dependent variable was the locus of control pattern of the teacher. The method used to collect the data was a simulation device termed Decisions in Teaching. The color motion picture, "Don't Tell Me, I'll Find Out," was used as the stimulus. At nine points during the film, the projector was stopped and the teachers responded with their agreement or disagreement to six possible decisions of what could occur next in the classroom. The response was based on a 5-point Likert scale ranging from complete agreement to complete disagreement on each of the nine scenes, two were representative of each of the three loci of control. Three grand totals (one for each locus of control) were used to describe a profile of the teachers' predicted behaviors if they were to use materials like those in the film. The predicted patterns of locus of control as measured by the Decisions in Teaching simulation device was used as the dependent measure in each of the six studies. Except where noted, a multi-variate analysis of variance was employed and simultaneous T contrasts were used as follow-up tests where the multivariate F was significant. In all of the studies in this investigation, the curriculum context was the Science Curriculum Improvement Study (SCIS) elementary science program.

Study One. Thirteen teachers were used to test the hypothesis that no differences existed between the predicted behaviors and observed behaviors of teachers. Predicted behaviors refers to what the teacher thought was most likely to occur next when viewing the film. Observed behaviors refers to what the teacher did in the classroom when and if an instance similar to one of those on the film actually happened. Each of the 13 teachers had taught SCIS for at least one year. The teachers viewed and responded to the stimulus film prior to beginning the school year. During the next six months, two trained observers recorded observations of the teachers at least twice a week. A Chi square test was employed to test the hypothesis.

Study Two. This study was designed to assess any change in pattern of locus of control after two-week or four-week workshop training in SCIS techniques. Seventy-six teachers in three different geographical locations responded to the Decisions in Teaching film before and after the SCIS training workshops. A one-group pretest-posttest design (Campbell and Stanley, 1966) was used.

Study Three. The purpose of this study was to determine if there were differences in locus of control responses between teachers who elected to attend SCIS workshops and those who elected not to attend such workshops. Sixty-nine teachers beginning SCIS workshops were compared to 51 teachers from the same schools who were not attending SCIS workshops. This was a comparison of nonequivalent groups prior to any intervention or treatment.

Study Four. One hundred twenty teachers were involved to compare locus of control patterns for teachers using book-centered science and teachers using activity-based programs.

Study Five. This study was designed to compare teachers just finishing SCIS workshops and those having taught SCIS for one or more years to determine if a regression in pattern of behavior occurred. The thought of the researcher was that persons trained to use the new materials might initially adopt their philosophy but slowly return to old ideas as time passed.

Study Six. Eighteen SCIS training staff members were compared to teachers using the SCIS program for one or more years to determine if teachers' responses to the Decisions in Teaching instrument were different for those of the SCIS staff. The analysis employed a rank ordering of each of the nine scenes in the film from most student oriented to least student oriented. These rankings were then compared.

Findings

Study One. Chi square test showed no significant differences between predicted behaviors and observed behaviors. Classroom observers noted 152 situations in the classrooms similar to situations seen in the film. Of those 152 situations, 113 showed behavior matching predicted behavior.

Study Two. A significant multivariate F was obtained. Simultaneous T2 contrasts revealed that the difference was only on the teacher-oriented score. Teachers who completed the two- and four-week workshops scored between 4 and 13 points lower on the teacher-oriented score than prior to the workshop. In other words, the SCIS trained teachers were less likely to expect teacher-oriented actions in the classroom than were teachers not in the workshops.

Study Three. No significant differences were found between teachers electing to attend SCIS workshops and teachers electing not to attend SCIS workshops on any of the locus of control scores. The researcher concludes that teachers volunteering to attend SCIS workshops were not more disposed to the SCIS philosophy than teachers who had not volunteered.

Study Four. The MANOVA results indicate a significant difference between teachers using activity-oriented science and teachers using book-centered science. Again the difference was only on the teacher-oriented measure with teachers using activity-centered science agreeing with fewer teacher-oriented behaviors than the teachers using book-centered science.

Study Five. No significant differences were noted between teachers just completing SCIS workshops and those having taught SCIS for one or more years. In other words, the philosophy of teachers toward locus of control in the classroom was not different for newly trained and veteran teachers.

Study Six. Results indicate that experienced SCIS teachers and SCIS staff members could not be differentiated on the basis of their rank orderings of the nine scenes in the film.

Interpretations

Several conclusions were reached by the researcher. (1) The Decisions in Teaching simulation could be used to predict teacher behavior. (2) Involvement with curricular materials that are activity-oriented results in at least prediction of teacher behavior, if not teacher behavior itself, that is less teacher oriented. (3) The Decisions in Teaching simulation can differentiate between teachers using book-centered science and teachers using activity-centered science.

The researcher also concludes that the evidence supports the assumption that "quality curricular materials would change teacher behavior"; that a wide variety of teaching styles may be acceptable for beginning new programs, and that "with adequate training and/or experience, teachers can discern the locus of control which is consistent with that of curriculum designers."

ABSTRACTORS' ANALYSIS

The investigation appears to be a well conceived and conducted study, but the written report is somewhat confusing. Subheadings are used to denote each of the six studies, but the discussion sometimes shifts from that of a particular study to one on the investigation as a whole. Reorganization of the report or the use of additional subheadings would aid the reader in interpretation of the study.

The model of locus of control as defined by the researcher as the framework for analysis of classroom interactions seems reasonable. It should be noted that this use of the term locus of control is not the same as the more commonly used one (Rotter, 1966 and Rowe, 1978). The researcher operationally defines his use of the concept quite adequately.

The use of a simulation device to predict teacher behavior is an interesting one. The researcher has gathered evidence to indicate that actual behavior is associated with responses predicted in a simulation situation. Work by Butts and Dillashaw (1980) also indicates that actual teaching behavior can be predicted by simulation exercises. The description of administration of the simulation is clear. However, the procedure by which the classroom behaviors were selected and classified is not. We are told that 152 situations similar to those in the film were observed and that 74 percent of these matched the teachers' predicted responses. But how were classroom events judged to be similar to the film events and how were the teacher responses to them categorized? Thus questions relating to both the validity and reliability of the classroom observation measure are unanswered.

In study two, the possibility of pretest sensitization must be considered since the post-training exercise is conducted only two or four weeks later. It is not clear if all 76 teachers in the sample took both the pretraining and post-training exercise. In study six the researcher does not report a statistical test used to compare the rankings of the SCIS staff with SCIS teachers.

We question whether the assumption that "quality curricular materials would change teacher behavior" was actually an assumption underlying the development of programs. The researcher concludes that his evidence supports this assumption, but the investigation had more to do with training in use of curricular materials as a means of changing behavior. No evidence is given that use of the materials alone changed teachers' beliefs about control patterns in the classroom.

The conclusion that a wide variety of teaching styles may be acceptable for beginning new curriculum programs is likewise not justified by the evidence. It seems that a more appropriate conclusion would be that a variety of teaching styles may be acceptable if appropriate training is supplied as this investigation dealt heavily with training in SCIS materials. The researcher's conclusion that with training teachers can discern the locus of control of a curriculum program is supported by the evidence presented.

Aside from the conclusions themselves, a noteworthy part of this report is how the researcher deals with a problem accompanying use of pre-experimental designs. Nonequivalent groups are compared in Studies 3, 4, 5 and 6 in the report. The threats to the validity of findings from such studies are well known (Campbell and Stanley, 1966). To allay concerns about such factors as selection or mortality, the researcher needs to provide information to the reader so that a judgment can be made about their seriousness. There is no intention here whatsoever to suggest that the researcher has chosen inappropriate designs. Investigators working with inservice teachers rarely have the luxury of random assignment and the use of true experimental designs. Their choice then is for the best design under the circumstances. Since these designs are likely to be ones that allow alternative interpretations of findings, the researcher needs to deal with these possibilities.

Study three in this report is an example of a study done to answer a question relating to the threat of selection bias. Study two had shown that teachers enrolled in a SCIS workshop significantly changed their responses about locus of control. But the researcher says that perhaps these volunteers for training were predisposed to a change in philosophy. If this were so, one would need to examine teachers who did and did not volunteer for workshops to see if they responded differently to the locus of control instrument. Study three showed that they did not. The researcher has therefore shown that volunteer teachers are not different from their peers with regard to locus of control philosophy. This lessens the concern that there may have been a selection bias in Study 2.

This study is an important contribution to the field of teacher training in the area of teacher perception of behaviors appropriate for a given curriculum program.

REFERENCES

Butts, D. P. and F. G. Dillashaw. "The Problem Solving in Science Teaching Test (PST)." Science Education REports. Athens: University of Georgia, 1980.

Campbell, D. T. and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1966.

Rotter, J. B. "Generalized Expectancies for Internal Versus External Control of Reinforcement." Psychological Monographs: General and Applied 80(1), whole no. 609, 1966.

Rowe, M. B. Teaching Science as Continuous Inquiry. New York: McGraw Hill, 1978.

IN RESPONSE TO THE ANALYSIS OF

Berger, C. F. "Investigation of Teacher Behavior: Interaction with New Curriculum Materials," by F. Gerald Dillashaw and James R. Okey. Investigations in Science Education, 8(1): 49-54, 1982.

by

Carl F. Berger
The University of Michigan

I believe that Dillashaw and Okey have done an excellent job of identifying and abstracting the major points in the study and I do not have any comments about their abstract of the study. I do, however, have several comments about their analysis. Not all are to debate their interpretations; in fact, some are to clarify and to respond to the questions they have raised.

The abstractors' first comment, that the discussion shifts from that of a particular study to one of the investigation as a whole is well taken. The study was part of an overall long-term investigation done when the author was a member of the SCIS staff which attempted to answer questions raised by the staff regarding changes in teacher behavior. So that while each one of the studies could be thought of as an individual entity, they nevertheless fit as a coordinated whole. It is most difficult to write the paper as if they are six independent studies when in reality they do interlock quite heavily. Breaking the total investigation into a series of studies was an attempt, similar to that suggested by the abstractors, to utilize a subheading format.

The questions the abstractors raised as to the validity and reliability of classroom observation measures is well taken and is crucial. This study could have been thought of as an interesting exercise in predicted teacher behavior, but would have little validity if we did not observe those predicted behaviors in actual teaching. Thus, two observers attended every session of science teaching across grades 1-6 for 13 elementary school teachers for an entire year. This was an extensive study as part of a follow-up study to a cooperative college/school science NSF program and afforded us the opportunity to reduce or eliminate change in teaching behavior when an observer may be present in very few situations in a class. The observation techniques were done as follows. A checklist was made, using the items from the simulation and the observers merely checked when they saw the teacher perform items in the classroom that were on the simulation. It is hard to imagine how a more valid device could be constructed since it included every situation and response from the simulation measure as well as establishing reliability by observing teachers over an entire nine-month teaching period.

The above remarks would have been helpful to have included in the research design and procedure, but for the sake of space, they were left out. In

Study 2, the possibility of pre-test sensitization was considered early in the study as there were some teachers who were not available for pre-test but were available for post-test. Analysis indicated no statistical difference in post-test scores between those who had the pretest and those who had not had the pre-test in either two- or four-week training sessions. These results were not included in the study since there were so few teachers who had missed the pre-test administration.

In Study 6, correlations were made using the order or preference of statements on each situation with the reordering of the statements by the judges. The situations were then ranked in order from the most student-oriented to the least student-oriented and these rankings were then compared across groups. Rankings of eighteen staff members of SCIS compared to the post-institute and post-institute plus extra years of service indicated no difference in the rankings of the responses. Thus, no statistical test was done since there was one-to-one correspondence between the correlations of most student-oriented to least student-oriented preferences.

The author must agree with the reviewers that the study dealt heavily with the training and use of curricular materials as a means of changing behavior. One can conclude, however, that if training only were responsible for the effect we would see a regression to the mean occurring after the training had been completed and teachers used the materials. In fact, teachers continued to respond similar to that of the SCIS staff and did not revert back to preferences they held prior to the study. This is in direct opposition to some recent studies of other training sessions dealing with ISCS and similar science curricular projects in which strong regression to the mean was noted after teachers left a training situation and were faced with the reality of their own teaching situation. One, therefore, can argue that quality curricular materials can change teacher behavior, but it may be necessary to start with a training session rather than just starting with quality curricular materials. The reviewers are quite right in pointing out this omission.

The conclusion that a wide variety of teaching styles may be acceptable for beginning new curricular programs was based upon the observation that no single locus control of particular response preference was given by any one teacher for all situations. It appears that the responses are very situation-specific and while teachers may tend to be more student-oriented using the SCIS curriculum materials and participating in the SCIS training, specific situations are not necessarily always answered by a student-oriented locus of control. This does not diminish the reviewers' comments that a variety of teaching styles may be acceptable if training is supplied with this investigation, but the research was concerned that a seemingly student-oriented curricular program not produce only student-oriented predictions of behavior by teachers.

The abstractors are quite correct in noting that investigators working with inservice teachers do not have the luxury of random assignment. In addition to the techniques with which the reviewers have noted for coping with the problem of a nonexperimental design, the researcher obtained training sessions in quite diverse geographic locations as well as quite diverse training styles. Using the West Coast, the Midwest, the South, and the Northeast as locations for gathering data with the concomitant differences in styles of training, the researcher hoped to reduce problems inherent in nonexperimental design.

As might be expected, the researcher is pleased that the reviewers have found the study to be an important contribution to the field of teacher training. Since this study was completed, over 2,000 teachers have participated in the use of the device as both a research tool and as a tool for inservice teacher training. Such instruments in practical situations can add even further to our knowledge of teacher education training situations and new curricular materials.

Renner, J. W. "The Relationships Between Intellectual Development and Written Responses to Science Questions." Journal of Research in Science Teaching, 16(4): 279-299, 1979.

Descriptors--Cognitive Measurement; Educational Assessment; *Educational Research; *Evaluation Methods; *Learning Theories; Measurement Techniques; *Questioning Techniques; Science Education; Science Tests; Secondary Education; *Secondary School Science; Written Language

Expanded abstract and analysis prepared especially for I.S.E. by M. E. Miller and M. C. Linn, University of California.

Purpose

The research under review was conducted by John W. Renner of the University of Oklahoma and members of the Cognitive Analysis Project (CAP). Its purpose was to assess the level of intellectual development (concrete or formal operational) of a group of high school students by examining their written responses to science questions. It was hoped that by evaluating a number of such responses, results comparable to those produced by a standard Piagetian task-interview could be obtained.

Renner justifies the use of a written-response format in the stated hypothesis of the study, namely:

> that examining the use persons make of language in explaining phenomena would reveal their logic structures. Said another way: since language is based upon the use of logic, examining the use of language reveals logic.

Rationale

Within Piaget's theory of intellectual development, the attainment of formal operations marks the emergence of fully mature logical structures, usually occurring sometime between 13 and 16 years of age. Included within the stage of formal operations are a number of formal schemes, such as combinatorial and proportional reasoning. The degree to which these schemes are present is typically assessed by means of a one-to-one interview employing a number of laboratory-type tasks (bending rods, balance beam, etc.) first introduced by Inhelder and Piaget (1958). A typical Piagetian interview takes from 30 minutes to one hour to administer, and is inappropriate for evaluating large groups of subjects. A paper-and-pencil test which reliably measures formal reasoning ability would therefore serve an extremely useful diagnostic function.

In the present research, the investigator has assumed that the use a subject makes of written language will closely parallel his reasoning processes. He further assumes that knowledge of students' reasoning abilities will aid teachers in planning educational programs.

Research Design and Procedures

Trained interviewers administered four tasks (conservation of volume, control of variables, balance beam, combinations of colorless liquids) to each of 297 tenth, eleventh and twelfth-grade subjects. Each subject's performance was awarded points as follows:

LEVEL IIA	- EARLY CONCRETE	(1 POINT)
LEVEL IIB	- CONCRETE	(2 POINTS)
LEVEL IIIA	- EARLY FORMAL	(3 POINTS)
LEVEL IIIB	- FORMAL	(4 POINTS)

Individual task scores were then summed to provide an overall assessment of a subject's developmental level. Cumulative scores were scaled as follows:

(4 - 8) = CONCRETE
(9 - 11) = TRANSITIONAL
(12 - 15) = FORMAL

These are the scores which CAP attempted to predict using subjects' written responses to science questions.

Written questions, called "incidents," were developed by members of CAP. These questions required subjects to think "scientifically" but were intended to require no special scientific knowledge. The reading level of each incident was controlled. Subject responses were used to generate an ordinal response scale for each incident. Scores from the incidents were then entered into a regression equation which was used to predict the cumulative interview scores.

In addition, 143 of the subjects were also given the Embedded Figures Test (EFT), a measure of the field dependence-independence construct. These scores were used to improve the predictive power of the regression equation.

Subjects were drawn from three high schools in Oklahoma. Subject-selection procedures are not specified by the author, and cannot be assumed to be random.

Findings

The multiple correlation between the four CAP incidents and the cumulative Piagetian interview scores was R = 0.62 (SE = 2.04), accounting for 36 percent of the interview-score variance. When the EFT scores for 143 subjects were used in addition to their incident scores, the obtained correlation was R = 0.70 (SE = 1.85), accounting for 49 percent of the variance. These same values, however, were obtained from an equation which used only three of the incident scores, but which retained the EFT. No simple correlations were reported.

Interpretations

Renner discusses two possible causes for the failure of the CAP incidents to achieve greater predictive power. The first of these is that the Piagetian interviews themselves are, of course, less than completely reliable. The second, and more important, reason considered is that some element included in the interviews is missing from the incidents. It is the lack of social interaction, Renner feels, which prevents the incidents from obtaining a higher multiple correlation with the cumulative interview scores.

ABSTRACTOR'S ANALYSIS

In the final paragraph of this article Renner says, "When a teacher knows the intellectual capabilities of the members of a class, decisions can be made about the types of concepts--concrete, formal, or both--which can be taught to that class." This statement suggests a relationship between concrete and formal reasoning as measured by Piagetian tasks and performance in learning situations. No conclusive evidence for this relationship exists.

Renner cites no evidence whatsoever for the relationship between performance on these tests and ability to learn in the classroom. On the contrary, the evidence reported by Renner suggests that the correlations between interviews and group measures of concrete and formal thought are low. Only 36 percent of the variance in the individual interviews is accounted for by the group tests. It seems irresponsible to recommend that teachers make decisions on the basis of these tests when they have such poor reliability. Furthermore, even if the tests were completely reliable, Renner gives no justification for using them to decide upon the learning activities for individuals in a class.

A review of the research on group-administered measures of cognitive development, with particular regard to the adequacy of written-response evaluations, would have served a useful orienting function and helped the reader to evaluate the present research. No such review of the literature is provided.

Two problems inherent in paper-and-pencil assessments of cognitive development should be pointed out: a) the activities of the subjects cannot be observed, and b) their responses cannot be proved. In the present research, the use of written responses risks loss of validity by confusing logical reasoning with writing skill, thus adding to the difficulty of inferring formal reasoning ability. Even if language does reflect reasoning, it is naive to suppose that adolescent subjects express themselves equally well verbally and in writing. Written language skills usually lag well behind verbal skills, and writing about one's own thought processes necessarily provides only a dim reflection of the reasoning of many subjects.

In addition, group tests are likely to be less reliable than interviews. By probing a subject regarding his performance on a task, the interviewer is able to arrive at a fairly accurate understanding of that subject's reasoning. Only then can he say that a subject has attained a certain level of reasoning in a particular logical domain. Group tests, of course, do not permit probing.

Renner hypothesizes that "examining the use of language reveals logic." This is based upon an interpretation of Piaget which he does not adequately support. The fact that logical behavior precedes language, which Renner cites in support of his method, does not imply that linguistic structures directly reflect logical structures. According to Piaget (1977):

> operatory structures constitute, even if their elaboration is based on verbal behavior, relatively complex systems not included as systems in language itself (p. 120)

Several questions arise concerning the instruments employed by the CAP staff:

1. Even though the reading level of each incident was controlled, the language used is not specific with regard to the expected performance. For example, in the separation of variables task (the Geranium problem), students were told to "describe the experiments you need to do in order to test whether or not each of these factors is important to the growth of geranium plants." While it is probable that most high school students have the word "experiment" in their vocabularies, it is much less certain that they construe it to mean a controlled,

scientifically valid experiment of the type the investigators intended. If a subject then fails to perform a series of controlled experiments involving the several variables under investigation, it does not follow that he is necessarily incapable of doing so.

2. Although it is reported that members of CAP "evaluated the incidents to determine if the complete solution to the problem required formal thought," the method by which such evaluations were made is not reported. In at least one instance, there appears to be a lack of correspondence between the incident and the ability it purports to measure. The Rock and Scale incident, which was designed to assess combinatorial reasoning, neither resembles other measures of combinatorial logic (Inhelder and Piaget, 1958; De Luca, 1978), nor does it appear to require combinatorial reasoning for its solution.

3. The response scales for the CAP incidents are ordinal in nature, but certain higher-level responses are neither unambiguously better than other lower-level responses, nor are they necessitated by the incidents themselves. For instance, level-5 responses to the Rock and Scale problem appear to be neither more adequate than level-4 responses nor logically required by the question.

4. Although Renner indicates that there was considerable difficulty in constructing the incidents, he does not report their reliability. It is impossible to tell, therefore, to what extent the abilities required to solve these problems are related to one another or, conversely, to what extent each of them involves a unique ability component.

5. The rationale for including the EFT is not reported within the context of the original research design. It is unclear whether this measure was an integral part of the study or if it was included only at a later time. It is quite possible that the EFT is measuring only general intellectual ability. While the predictive power of the regression equation was improved by the inclusion of EFT scores, these results are left unanalyzed.

6. Renner does not discuss the problems involved in developing the Piagetian interviews. As discussed by Linn (1977), there are many difficulties involved in translating the task descriptions from Inhelder and Piaget into actual interviews. As it is, it is impossible

to know exactly what was measured by the interviews employed by CAP. Interview reliability could easily have been estimated by calculating alphas for the four items, which would at least have allowed the reader to know whether each of the interviews was measuring the same ability.

The major finding of Renner that the incidents were not highly correlated with the interviews could be explained by a large number of factors including unreliability of both the incidents and the interviews. It is clear, however, that some variance in the interviews is not represented in the incidents since the addition of EFT scores to the regression equation accounts for more variance than was accounted for by the incidents alone.

The questions raised concerning the relationships among the incidents and interviews and EFT scores could have been answered more fully by a more complete use of correlational analysis.

First, correlations between the incidents and the separate interviews would greatly aid interpretation. We are unable to tell, for instance, whether the Geranium incident (separation of variables) is positively correlated with the separation of variables interview (bending rods). In addition, simple correlations between each of the items would enable the reader to determine whether the various measures of proportional reasoning, for example, are more related to one another than to the measures of other abilities.

Second, Renner's use of multiple regressions is arbitrary; that is, it would be just as reasonable to use Piagetian scores to predict incident scores as to use incident scores to predict Piagetian scores. It is especially important that the correlation matrix used to generate the regression analysis be available to enable the reader to understand the relationships in the data. Also, the multiple R is increased by the inclusion of the EFT scores, but because of the way the data are reported it is impossible to determine the extent of the overlap between the EFT and the various incidents. Furthermore, the regression weights of the items in the equation differ depending on when they are entered.

There is no discussion of the order in which variables are entered into the regression analysis.

There also appears to be some confusion concerning the appropriate technique for establishing interrater reliability. Interrater reliability cannot be computed for the Piagetian interviews because each interviewer tested different subjects. However, if interviewers were randomly assigned to subjects, it would be possible to see whether a main effect for interviewer could be observed. This is done by using a simple analysis of variance if a single task is employed, or a repeated measures ANOVA for multiple tasks. Renner uses a variant of this approach, but he reduces the sample size to 37 subjects per interviewer. He misinterprets the recommendation of Pearson and Hartley (1951) in thinking that only 37 subjects per interviewer should be used. In reality, using all the subjects would be the best way to determine whether there was a main effect for interviewer. Reducing the sample size merely reduces the likelihood of detecting an effect if one exists. As the probability of a main effect for interviewer was $p = 0.10$ using the reduced sample, it is quite likely that using the entire sample would have resulted in a different interpretation.

Finally, Renner uses analytic procedures which require interval scales. The justification given--namely, that there is no evidence that these are not interval-level data--is inappropriate. Also, using summed scores for the interviews would be more appropriate if each interview was standardized first.

Efforts to measure logical reasoning using group tests need to be guided by educational concerns. Renner states that teachers would like to know about the intellectual development of their students; however, it is not clear that such information would, in fact, be helpful to teachers. No relationships between performance on tests such as Renner employs and classroom performance have been established. It may well be that the tests employed by CAP simply measure the same things that are measured by achievement tests or intelligence tests. If this is the case, there would certainly be no justification for subjecting students to additional tests for which there is no apparent use. The tasks employed by Renner all come from the science domain; many come from physics. The abilities they purport to measure may or may not be measured similarly if the tasks were chosen from other disciplines or from naturally occurring situations. These matters deserve careful scrutiny before we recommend that teachers use such tests for assessment purposes or in the planning of educational programs.

REFERENCES

DeLuca, Frederick T. "Applications and Analysis of an Electronic Equivalent of Piaget's First Chemical Experiment." Journal of Research in Science Teaching, 16(1): 1-11, 1979.

Inhelder, B. and J. Piaget. The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books, 1958.

Linn, M. C. "Scientific Reasoning: Influences on Task Performance and Response Categorization." Science Education, 61(3): 357-369, 1977.

Piaget, J. The Child and Reality. New York: Penguin Books, 1977.

Pearson, E. S. and H. O. Hartley. "Charts of the Power Function for Analysis of Variance Tests Derived from the Non-central F Distribution." Biometrika, 38: 112-130, 1951.

IN RESPONSE TO THE ANALYSIS OF

Renner, John W. "The Relationship Between Intellectual Development and Written Responses to Science Questions" by M. E. Miller and M.C. Linn. Investigations in Science Education, 8(1): 60-68, 1982.

by

John W. Renner
Professor of Science Education
University of Oklahoma
and
Larry E. Toothaker
Professor of Psychology
University of Oklahoma

In reviewing "The Relationships Between Intellectual Development and Written Responses to Science Questions," Miller and Linn assigned it a purpose that represented the procedures used to gather data and not the purpose for which the research was done. The reviewers state, "Its purpose was to assess the level of intellectual development (concrete or formal) of a group of high school students by examining their written responses to science questions." The article states (p. 279), "...the Cognitive Analysis Project was conducted to design techniques (if possible) which could be used to collect written information from everyone in an entire group simultaneously which would allow judgements to be made about the intellectual development of each individual in the group." The reviewers make the project appear to be a measurement project where, incidentally, science questions are used. In reality the purpose of the Cognitive Analysis Project (CAP) was to design and validate techniques which could be used to do what reviewers state was the purpose of the CAP. The difference in what the reviewers preceived and reality is an important one; one is developmental research, the other is a status study.

The CAP hypothesized that the use persons make of language could be used to evaluate their use of logic and, consequently, their level of intellectual development. The reviewers quoted the hypothesis of the CAP early in their review. A few paragraphs later, however, this statement is found "...the investigator has assumed that the use a subject makes of written language will closely parallel his reasoning processes." Unless the reviewers do not accept that a hypothesis cannot be false, seeing how they arrived at their conclusion that the CAP assumed the hypothesized relationship is difficult. No one in the CAP made such an assumption. Our purpose was to design techniques to study the relationships between language and logic if they existed. The degree of such relationships was determined by the correlation coefficients found. The CAP most certainly did hypothesize such a relationship which the reviewers quoted. Seeing why the reviewers accused the author of the foregoing assumption is difficult.

The reviewers also state, "He (the author) assumes the knowledge of students' reasoning abilities will aid teachers in planning educational programs." Two research studies have demonstrated that concepts can be evaluated as requiring concrete or formal operational thought (Lawson and Renner, 1975; Cantu and Herron, 1978). Furthermore, those research studies have stressed the importance of knowing both the type of reasoning required to understand a concept and the levels of reasoning of the students. Cantu and Herron say (1978, p. 141), "...much of what we teach in science appears to require formal operational thought..." That statement seems to say that teachers need to know how to evaluate content as to the type of thought required to understand it. Those authors also state (Cantu and Herron, 1978, p. 14), "...many students who enroll in science do not use formal-operational thought." If a teacher is going to teach that content which matches the intellectual level of the students, the assumption "knowledge of student's reasoning abilities will aid teachers in planning educational programs" seems warranted. The validity of the assumption is further supported when the finding of both of the foregoing research studies that concrete-operational students have little or no success with formal-operational concepts is considered.

In the "Abstractor's Analysis" section of the Miller and Linn review, they quote the last sentence of the article which is, "When a teacher knows the intellectual capabilities of the members of a class, decisions can be made about the types of concepts--concrete, formal or both--which can be taught to that class." In view of the two research studies just cited, that statement--it did when it was written and still does--seems reasonable. Miller and Linn say, "This statement suggests a relationship between concrete and formal reasoning as measured by Piagetian tasks and performance in learning situations. No conclusive evidence for this relationship exists." The research studies just cited--Lawson and Renner, 1975, and Cantu and Herron, 1978--were separated chronologically between intellectual development and "performance in learning situations." In the Lawson-Renner study interviews with Inhelder-Piaget tasks were conducted, and Cantu and Herron used the Longeot test to measure intellectual development. The criticism of Miller and Linn, therefore, seems unfounded. (The assumption is made that the reviewers were aware of the two research studies cited here.) Miller and Linn did state that "No conclusive evidence..." exists for the final sentence in the article. Could these reviewers be stating that they do not accept the research cited here as "conclusive" enough for them? If so, why did the reviewers not state that, while some evidence exists to support the article's final statement, it is not "conclusive"?

Miller and Linn state, "Renner cites no evidence whatsoever for the relationship between performace on these tests and ability to learn in the classroom." The Lawson-Renner 1975 study was cited (p. 280). The Cantu-Herron article was not cited. There is a chronological reason for the omission. The research for the CAP was done in 1976; the report was prepared before the Cantu-Herron article appeared. The word "whatsoever" in the Miller-Linn statement suggests that the reviewers missed the inclusion of the Lawson-Renner work (p. 280). In view of what has just been said, the harsh judgment of Miller and Linn of the article's author hardly seems warranted.

These reviewers did not, however, stop their condemnation of the research with the statement just cited. The just-cited statement is followed with, "On the contrary, the evidence reported by Renner suggests that the correlation between interviews and group measures of concrete and formal thought are low." Notice that the correlation coefficients (arrived at in the CAP by using a regression equation) are not cited at this point in the review. Depending upon the arrangement of factors used in the regression equation, those correlation coefficients are 0.62 and 0.70. (Those correlation coefficients had been cited earlier in the review.) If the reviewers believe correlation of 0.62 and 0.70 are low, perhaps they have the responsiblity to explain what they would have found acceptable.

The indictment of the research for low correlation is followed by, "Only 36 percent of the variance in the individual interviews is accounted for by group tests." In fact, 38 percent and 49 percent--depending upon the arrangement of elements used--of the variance referred to is accounted for by the group tests. (If, for example, the correlation had been 0.80--a large correlation--only 64 percent of the variance would have been accounted for.) Miller and Linn neglected to point out that one arrangement of elements in the regression equation raised the variance accounted for to 49 percent.

The 36 percent of the variance statement led Miller and Linn to state, "It seems irresponsible to recommend that teachers make decisions on the basis of these tests when they have such poor reliability." Miller and Linn are, it must be assumed, saying that correlation coefficients of 0.62 and 0.70 between written measures of intellectual development and the Piagetian-type interview represent "poor reliability." Perhaps that is a judgment a potential user of the written tasks should make. It seems that that judgment can be made by a potential user only by examining the hypothesis of the research and deciding if correlation coefficients of 0.62 and 0.70 are sufficient for the user's purpose. To give future users of the tasks the data from the research and assume they will make the judgments hardly seems "irresponsible." (A question could be raised about the reviewers not quoting the correlation coefficients at the point in the review the research was condemned and computing the percentage of variance accounted for on the lower of the two coefficients.)

The condemnation of the value of the research to education by Miller and Linn does not stop with the foregoing quotations. They continue, "Furthermore, even if the tests were completely reliable, Renner gives no justification for using them to decide upon the learning activities for individuals in a class." In order to see how the research is useful to teachers, some analytical thinking is required. On page 280 of the article these sentences are found, "Formal concepts are not understood by those reasoning concretely." Research (Lawson and Renner, 1975) supports the foregoing. Now in order to use those statements as justification for classroom use of the science tasks developed in the research, one has to reason that, since formal concepts are not understood by concrete learners, the teacher must understand how to identify concrete learners and formal concepts. The science tasks developed by the CAP (which correlate with the Piagetian interviews at

0.62 or 0.70) help teachers identify concrete learners. While the assumption of the article's author that such a train of analytical thinking would take place seems to be optimistic, it hardly seems "irresponsible."

Miller and Linn state that "A review of the research on group-administered measures of cognitive development--would have served a useful orienting function...." They are no doubt correct. There are, however, several other factors that would have been helpful to include, but the article ran 21 journal pages just to describe the procedures used and give the results. A conscious decision was made not to include such a literature review. Since the article was published in a refereed journal the omission of such a review apparently troubled Miller and Linn more than it did the referees.

Several paragraphs of the review are devoted to the inherent difficulties with paper-and-pencil assessments of intellectual development. The reviewers point out that the reasons for the difficulties stem from the fact that the activities of the subjects cannot be observed and their responses cannot be probed. Miller and Linn present their conclusions about the difficulty of written assessments of intellectual development as if they are new contributions to judging the value of the results of the CAP. They make no mention that the following statement was included in the article (p. 298), "When assessing the presence or absence of those major intellectual structures with any instrument that does not allow for immediate feedback and two-way communication, the element that social transmission contributes to the rating the student receives is neglected. In other words, of the two scores being correlated (the interview score and the score on the written tasks), one contains the element of social transmission and the other does not." While the words "probed" and "observed" were not used in the original article, the fact that "feedback," "two-way communication," and social transmission" are suitable synonyms seems evident. As further evidence that Miller and Linn seem intent on ignoring what was said in the published report of the CAP, consider this sentence from the review, "...group tests are likely to be less reliable than interviews." The published report says (p. 299), "The writer hypothesizes that removing the element of social transmission from the process of determining what a particular student's intellectual level is reduces the validity of the process. If, of course, the validity of the process is reduced, the reliablity (and the correlation) of different assessments of the same attribute would also be reduced."

Several statements in the Miller-Linn review raise questions regarding whether or not they understood the basic hypothesis of the CAP. Consider this statement, "Even if language does reflect reasoning, it is naive to suppose that adolescent subjects express themselves equally well verbally and in writing." Based on many years of classroom experience with adolescents, this writer agrees with Miller and Linn. But the CAP did not equate the type of language used in the interviews with the type of language used in responding to the incidents. The language of the interview was used to rate the students' responses on each Piagetian task and assign a rating--IIA, IIB, IIIA and IIIB--to those responses. The language of the interview was not considered again. What was considered was the type of language used by a particular level of student in responding in writing to a particular task. The fact that the students' written language lags "well behind" oral language is not relevant because at no time were the written and oral language of the same students considered simultaneously. That would have had to happen for the researchers to have been guilty of being "naive" as Miller and Linn charge. Again, the question must be raised about the understanding of the CAP Miller and Linn had.

The general hypothesis of the research was that "examining the use persons make of language in explaining phenomena would reveal their logic structures" (p. 281). Miller and Linn say, "This is based upon an interpretation of Piaget which he (Renner) does not support." The data leading to that hypothesis and the sources of those data are found on pages 280 and 281 in the article. The reader will have to judge whether or not those data are "adequate." Those data apparently are not adequate for Miller and Linn, but then they do not state what would be, from their frame of reference, adequate. The reviewers include a quotation from Piaget, but no attempt is made to explain how it explains that adequacy or inadequacy of the general hypothesis of the CAP.

In discussing the procedures used by the CAP, Miller and Linn make this statement, "Subject-selection procedures are not specified by the author, and cannot be assumed to be random." Whether or not the selection of the subjects was random is irrelevant. The CAP was not conducted to describe the intellectual development of Oklahoma secondary school students. If that had been the purpose of the CAP, then a random sample (probably stratified) would have been essential. The CAP interviewed each student with Piagetian tasks and computed his/her score. Each student interviewed completed the written tasks, and an analysis was made of the type of language used by students who earned a specific score. The fact that a student with a particular score was from a city, a rural area or a private school was irrelevant. The CAP was not interested in what types of schools foster what kinds of reasoning (that we have already done; Renner; et. al., 1976, Chapter 6), rather the CAP was interested in what types of written language students with particular Piagetian interview scores used in responding to the science tasks presented them. In that case, this writer contends whether or not the sample was random is irrelevant.

Miller and Linn raise six points about the instruments employed in the CAP. Each of those points will be commented upon and each comment here will bear the same number used in the review.

1. Miller and Linn argue that the language of the questions may not have been understood by the students responding to the science tasks. There is, of course, no way to support or refute the reviewers' contention. The scales constructed for evaluating the particular incident the reviewers center on and the relationship between performance on that incident and the Piagetian tasks seem to suggest their observation is not supported. Furthermore, the idea that the subject did not understand the language and if he/she had, the question would have been answered correctly is a criticism that has been leveled at the adminstration of all Piagetian tasks. The reader must decide if the reviewers' criticism has validity.

2. Miller and Linn criticize the fact that the report of the CAP did not report how the decision was made that each incident required formal thought. That is a legitimate criticism; the article did not report that procedure. That procedure is too lengthy to report here and, furthermore, this rebuttal to Miller and Linn is not the time to introduce new data. The reviewers criticize "The Rock and Scale" incident which was designed to measure combinatorial logic and to require the IIIB level of thought because it "neither resembles other measures of combinatorial logic..., nor does it appear to require combinatorial reasoning for its solution." Miller and Linn, however, do not include any explanation which illustrates why or how they arrived at their conclusions.

3. Miller and Linn criticize the response scales, center their attention on the scales for the "Rock and Scale" incident, and say, "For instance, level-5 responses to the Rock and Scale problem appear to be neither more adequate than level-4 responses not logically required by the question." Here, again, the reviewers give no explanation which justifies their conclusion. After considering a great many responses, the staff of the CAP disagrees with the opinion of Miller and Linn.

4. Miller and Linn criticize the article for not reporting the reliability of the incidents. Earlier the incidents were criticized for their "low reliability." Those two positions the reviewers have taken do not seem mutually supportive. How could the reviewers know the reliabilities were "low" if those reliabilities were not reported? If the reviews are referring to test-retest reliability of the individual incidents, they are correct. No such reliabilities were found, nor was it the intent of the CAP to do so. Our intent was to correlate performance on the incidents with performance on the Piagetian interviews.

5. Miller and Linn level the following criticism at the report of the CAP, "The rationale for including the EFT is not reported within the context of the orginal research design." This writer disagrees with that statement. If pages 294-295 of the article are consulted, the rationale for including the EFT in the design from the beginning of the research is indicated. Consideration of it as a valuable tool however, was dropped in the early days of the CAP after correlating performance on it by 412 students with performance by the same students on the Piagetian interview and receiving a Pearson r of 0.56. Later in the project the attention of the staff was returned to the EFT. The "quite possible" suggestions the critics make may in fact be true. Their suggestions for further analysis of the data from the CAP may have merit, but the staff of the CAP believe that such an analysis went beyond what the CAP was for.

6. Miller and Linn criticize the report of the CAP for not discussing "the problems involved in developing the Piagetian interviews. As discussed by Linn (1977) there are many difficulties involved." The published report of the CAP was taken from the complete 164 page report of the project (Renner, Pricket and Renner, 1977), and

> that report was referenced in the section of the published article (p. 281) where the interviewing procedures were discussed. The data reported included the fact that "Data on 919 interviews of four tasks each were being conducted." The complete interviewing protocols were not included in the published article because of the length of the article; they are included in the reference cited. This writer can only assume that the reviewers consulted the full report and knew the CAP did not intend to discuss the "difficulties involved."

Near the end of their review of the published report of the CAP, Miller and Linn again return to their insistence "that the incidents were not highly correlated with the interviews..." As was stated earlier a correlation of 0.62 between the interviews and the incidents exists and, if the EFT is added, that correlation raises to 0.70. A potential user of the research will have to decide if those correlation coefficients are adequate for the use he/she intends.

In the same paragraph as the sentence just quoted, Miller and Linn make this statement, "...the addition of EFT scores to the regression equation accounts for more variance than was accounted for by the incidents alone." As was stated earlier, the correlation between performance on the EFT and performance on the Piagetian interviews was 0.56. The highest correlation between performance on the Piagetian interviews and the most productive combination of incidents was 0.62. So if Miller and Linn are saying that the EFT alone is better as a predictor of how a student will score on the interview than are the incidents used in combination, they are incorrect. If, however, they are saying that adding the EFT performance of a student to that student's performance on the incidents increases the predictablity of the interview score (and reduces the variance), they are correct. Again the purpose of the CAP was to produce a written instrument that could be used to measure intellectual development. The findings are that adding the EFT score to the incidents score improves the written instrument; the advice of the CAP to potential users is, use it!

Miller and Linn criticize the article's author for not including the "Correlation between the incidents and the separate interviews..." That is a just criticism. The only defense that is offered is that the purpose of the CAP was to produce a wordable tool and the article's length would have been greatly increased if all such correlations and the mandatory accompanying discussion had been included.

Miller and Linn call attention to the fact that "Renner uses analytical procedures which require interval scales." They cite this as "inappropriate." The author's justification that there is not evidence that suggests they are not interval scales. Perhaps that is true. But Miller and Linn offer no alternative nor do they refute the author's contention. The critics also say, "Also, using summed scores for the interviews would be more appropriate if each interview was standardized first." The unspoken element in the Miller-Linn comment is (this writer believes) a thinly-disguised attack upon assigning a student one score from a entire interview. Consider this statement, "Adding scores from tasks such as these simply increases one's ability to reliably measure the extent to which the underlying formal operations have developed and how widely applicable they are if, in fact, they have developed" (Lawson, 1977). While evaluating what Miller and Linn mean by "more appropriate" is not possible, the foregoing quotation suggests that producing a single score for a single individual is not inappropriate.

The inclusion of the correlation matrix of PS, EFT, S, F, G, and R would have been helpful to the reader. However, the argument that "it is impossible to determine the extent of the overlap between the EFT and the various incidents" is included in the criticism of a lack of a correlation matrix. Also, the overlap between EFT and the incidents is unimportant since it is precisely because the EFT was entered separately that the multiple correlation increased from 0.62 to 0.70. This increase is not dependent upon the relationships between EFT and the various incidents but *in spite* of those relationships. The unique contribution of the EFT is given by the difference in the squared multiple Rs of 0.1056; that is, 10.56 percent of the variance in the PS score can be attributed to the EFT over and above that already contributed by the incident variables. The critics then state that "the regression weights of the items in the equation differ depending on when they are entered. There is no discussion of the order in which variables are entered into the regression analysis." For a given set of predictor variables, the regression weights are unique. Order of entry of variables is irrelevant. What is most likely meant by this statement is that the weights differ as a function of *how many* variables and *which* variables are used in the equation. Any order of entry of the incidents for a given equation will yield the same weights. Even the exact weights for the different equations are not important if you do not wish to measure the importance of these four incidents--and the CAP did not.

The critics of the research have, in our opinion, a questionable understanding of the power of a statistical test. The purpose in reducing the sample size when testing for interviewer effects was to prevent the extremely large total number of interviews per interviewer from permitting the F-test to detect (as significant) trivial differences between interviewers. Even the chosen sample size of 37 gives a power of 0.95 (a typographical error in the article) for differences of one standard deviation. That power (0.95) is large for *small* differences (one standard deviation) with N = 37. Using 155 to 253 cases would have resulted in power in excess of 0.99 for differences as small as 0.5 standard deviation and power of nearly 0.97 for 0.25 standard deviations. Indeed, it is quite likely that using all of the subjects would have led to a different conclusion; the question is, however, whether or not that conclusion would be correct. What the results of the N=37 analysis are telling us is that there are *not meaningful* differences in the interviewers, and, what the results of the proposed N = entire sample analysis would have told us is that there *are trivial* *or* *meaningless* differences in the interviewers. The sample of 37 was chosen by an intelligent, rational process to avoid detecting trivial differences in the interviewers. We wonder if the reviewers gave as much thought to their suggestion of using all the subjects.

The conclusion of the critical review of Miller and Linn contains many speculations which cannot be evaluated--possibly those speculations are true. There are two comments in that conclusion which must, however, be commented upon. Consider this statement, "The tasks employed by Renner all come from the science domain; many come from physics." Considering the title of the published article the fact that the tasks came from science should have been a surprise to no one. The writer cannot know what the reviewer's criterion reference for "many" is, but the fact is that, of the four science incidents isolated as useful in measuring intellectual development, two are from the physical science and two are from the biological sciences. Actually only one incident is specifically drawn from physics; one hardly seems like "many."

The second statement in the Miller-Linn conclusion that deserves comment is, "no relationships between performance on tests such as Renner employs and classroom performance have been established." That criticism was dealt with earlier in this rebuttal and, it is hoped, the point has been made that such relationships do exist. As was pointed out, however, some analytical thinking is necessary to understand those relationships.

REFERENCES

Cantu, Luis L. and J. Dudley Herron. "Concrete and Formal Piagetian Stages and Science Concept Attainment." Journal of Research in Science Teaching, 15(2): 135-143, 1978.

Lawson, Anton E. "Author's Response." Journal of Research in Science Teaching, 14(4): 365-369, 1977.

Lawson, Anton E. and John W. Renner. "Relationships of Science Subject Matter and Developmental Levels of Learners." Journal of Research in Science Teaching, 12(4): 347-358, 1975.

Renner, John W.; Dianna K. Prickett; and Michael J. Renner. Evaluating Intellecutal Development Using Written Responses to Selected Science Problems. University of Oklahoma, 1976.

Renner, John W.; Donald G. Stafford; Anton E. Lawson; Joe W. McKinnon; F. Elizabeth Friot; and Donald H. Kellogg. Research, Teaching, and Learning with the Piaget Model. Norman: University of Oklahoma Press, 1976.

IN RESPONSE TO THE ANALYSIS OF

Sunal, Dennis. "Analysis of Research on the Educational Uses of A Planetarium," by G. H. Krockover. Investigations in Science Education, 7(1): 57-59, 1981.

by

Dennis Sunal
West VirginiaUniversity

Though I agree with the abstractor that increasing the amount of information present describing the study would have made the study more clear, the article did describe enough of the important points to give those readers interested in the topic useful guidelines and/or sources for additional research. The extended abstract as written does not describe the purpose or nature of the research report as printed in volume 13, number 4, 1976, of the Journal of Research in Science Teaching.

The purpose was not to "analyze the development and use of a model..." as the abstract claims, but to analyze previous research data in terms of newly defined variables. As stated on page 345 in the JRST article, the report "concerns the development and use of a model for evaluating student outcomes involving a school-associated planetarium...and third, analysis of planetarium research studies in terms of the developed model." The abstract continues describing certain aspects of the study but selectively deletes or does not follow up areas which later are described as missing. Point by point, the abstract misses information on which six question are later asked.

Abstract 1. What was the basis used for the selection of the model?

Response: The abstract fails to note two paragraphs on the bottom of page 345 and one on the top of page 346 describing the origin and basis of model selection (cognitive, affective and process skill domains). This problem includes not reporting a cite for model development in a previous research study (Sunal, 1973).

Abstract 2. What specific procedure was used for using the model to analyze past research studies (p. 346)?

Response: On the next page (p. 347) of the article, two paragraphs under the heading of Procedure-Analysis describe the missing specific procedure.

Abstract 3. "Why did the author cover grades two through college (p. 346)?

Reponse: As stated on p. 346 in the article, all research studies dealing with planetarium education to date were used. The earliest study population researched to date was with second grade and the oldest, freshman college. As stated on p. 347, these studies were grouped and analyzed on separate levels--elementary, secondary and college.

Abstract 4. "What statistical analysis was used and why wasn't the level of significance reported" (p. 347)?

Response: Data from the previous research studies were secured and reanalyzed using "identical computer statistical analysis and significance levels of student data as performed in the original research reports" (p. 347). This is a problem. This information was originally included in the manuscript sent to JRST but was requested to be deleted due to taking up too much space. However, the reader, if interested, might contact the author or reveiw the individual studies, all cited in the references, for this information.

Abstract 5. "Why did the author use subgroup data when the original researchers used total score data for each student" (p. 347)?

Response: The researchers did not all analyze their data grouped into model categories-cognitive level, affective level and process skill area. Many used total scores from achievement tests. As described in the article, the author analyzed the questions on these tests and grouped them into these categories by levels--thus subgroups.

Abstract 6. "Why did the author use 15 variables" (p. 347)?

Response: This was described as the purpose in using a model to determine variables to measure. The author chose not to analyze one outcome, such as recall, or 100 but to analyze those areas which planetarium educators have reported as goals or objectives in using the planetarium (pp. 346-347). These resulted in the 15 variables.

In conclusion, the abstract as printed is confusing. Many of the issues cited should not be a problem to the mildly careful or interested reader. Problems may arise from the briefness of the report allowed by the editors of JRST. This not only involved the text of the article, but also four tables which the editors deleted from the final draft sent to the publishers. However, as a research report, the article's category, this will provide only some inconvenience in requiring some additional library check of references for those wishing to continue research in this line.